

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:16 ; Search time 31 Seconds  
(without alignments)  
34.387 Million cell updates/sec

Title: US-09-719-494-10

Perfect score: 36

Sequence: 1 SSIERRRL 8

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched:

908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq.101002:\*

- 1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:\*
- 2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:\*
- 3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:\*
- 4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:\*
- 5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:\*
- 6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:\*
- 7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:\*
- 8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:\*
- 9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:\*
- 10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:\*
- 11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:\*
- 12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:\*
- 13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:\*
- 14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:\*
- 15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:\*
- 16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:\*
- 17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:\*
- 18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:\*
- 19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:\*
- 20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:\*
- 21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:\*
- 22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:\*
- 23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	8	AA83942	MHC class I restr
2	36	100.0	8	AAV13407	HSV specific anti
3	36	100.0	8	AAV10350	T cell epitope/MHC
4	36	100.0	8	AA812542	GB-specific H-2b f
5	36	100.0	8	AAV67358	HSV glycoprotein B
6	36	100.0	19	AAE13452	Herpes simplex vir
7	36	100.0	19	AAE13453	Herpes simplex vir
8	36	100.0	22	AAU82071	Antigenic peptide
9	36	100.0	23	AAV13421	Conjugated control
10	36	100.0	26	AAV13420	Conjugated peptide

11	36	100.0	27	AAV13419	Conjugated peptide
12	36	100.0	30	AAE13454	Herpes simplex vir
13	36	100.0	40	AAU82093	T-cell specific bl
14	36	100.0	23	AAU82094	T-cell specific bl
15	36	100.0	694	AAE14666	Truncated HSVB po
16	36	100.0	795	AAV72062	HSV-2 strain SB5 C
17	36	100.0	854	AAV72113	HSV-2 strain SB5 C
18	36	100.0	6	AAV50312	Herpes simplex vir
19	36	100.0	903	AAV60244	Herpes simplex vir
20	36	100.0	903	AAV70426	Recombinant herpes
21	36	100.0	903	AAV71135	Herpes Simplex Vir
22	36	100.0	904	AAV80914	Sequence of Herpes
23	36	100.0	904	AAV14665	HSVB polyepitide
24	36	100.0	904	AAV41778	Glycoprotein B (9B
25	36	100.0	904	AAV41779	Glycoprotein B (9B
26	36	100.0	904	AAV00375	HSV-1 glycoprotein
27	36	100.0	904	AAV00376	HSV-2 glycoprotein
28	36	100.0	904	AAV34552	Herpes simplex vir
29	36	100.0	904	AAV34553	Herpes simplex vir
30	36	100.0	904	AAV72193	HSV-2 strain SB5 C
31	36	100.0	904	AAV74441	Herpes simplex vir
32	36	100.0	904	AAV74442	Herpes simplex vir
33	36	100.0	904	AAV17812	Herpes simplex vir
34	36	100.0	905	AAV80915	Sequence of Herpes
35	36	100.0	907	AAV71136	Herpes Simplex Vir
36	36	100.0	973	AAV70769	Glycoprotein B of
37	36	100.0	973	AAV14680	SVB surface antige
38	36	97.2	885	AAV92747	SAB virus gb glyco
39	35	97.2	891	AAV92746	B virus gb glycopr
40	35	97.2	891	AAV70293	Simian herpesvirus
41	32	88.9	8	AAV67359	HSV glycoprotein B
42	30	83.3	344	AAV76345	Yeast D-arabinose
43	30	83.3	372	AAV52570	Helicobacter pylori
44	30	83.3	465	AAV55734	H. pylori ORF 07ee
45	30	83.3	586	AAV00651	Aspergillus nidula

#### ALIGNMENTS

RESULT 1	AA83942	standard; peptide; 8 AA.
ID	AA83942	
XX	AA83942	
AC	AA83942	
XX	05-JUN-1996	(first entry)
DE	MHC class I restricted antigenic peptide #12.	
XX		
KW	MHC class I; antigen; MAGE; melanoma; breast cancer; bladder cancer;	
KW	Titermax; cytotoxic T-lymphocyte; tumour; pathogenic disease; bacteria;	
KW	parasite; human; animl.	
XX		
OS	Synthetic.	
XX		
PN	WO9528958-A1	
XX	02-NOV-1995	
PD		
XX	21-APR-1995	95WO-US04975.
PF		
XX	22-APR-1994	94US-0233496.
PR		
PA	(SLOK) SLOAN KETTERING INST CANCER RES.	
XX		
PI	Dyall R, Nikolic-Zugic J;	
XX		
DR	WPI; 1995-382848/49.	
XX		
PT	Cytotoxic T-cell induction by MHC class I-restricted peptide in	
PT	adjuvant - useful for treating tumours and bacterial or parasitic	
PT	pathogenic diseases	

XX Claim 11; Page 38; 50pp; English.  
PS  
XX The sequences given in AAR83931-49 are MHC class I restricted 8-12 amino  
CC acid antigenic peptides. This peptide is derived from Herpes simplex  
CC virus gB protein. These peptides may be administered to a subject  
CC in combination with a suitable adjuvant, pref. Titermax (RTM), to  
CC induce cytotoxic T-lymphocytes. This method may be used in the  
CC treatment of a tumour or a pathogenic disease, esp. diseases of  
CC bacterial or parasitic origin, in humans and animals, e.g. monkeys,  
CC dogs, cows, horses, etc.  
XX  
SQ Sequence 8 AA;  
Query Match 100.0%; Score 36; DB 16; Length 8;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SSIERFARL 8  
DB 1 SSIERFARL 8  
RESULT 2  
AAV13407  
ID AAV13407 standard; peptide; 8 AA.  
AC AAV13407;  
XX 19-JUL-1999 (first entry)  
DT  
DE HSV specific antigenic peptide B1 of glycoprotein B.  
XX  
XX Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;  
KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;  
KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;  
KW immune response; genetic immunisation.  
OS  
XX Herpes simplex virus type 1.  
XX  
XX WO9916710-A1.  
XX  
XX 08-APR-1999.  
PD  
XX 29-SEP-1998; 98WO-US20681.  
PF  
XX 30-SEP-1997; 97US-0060422.  
PR  
XX (CELS-) CEL-SCI CORP.  
PA (DYNE-) UNIV NORTHEASTERN OHIO.  
XX  
XX Rosenthal KS, Zimmerman DH;  
PI  
XX WPI; 1999-312418/26.  
DR  
XX  
XX New immunogenic conjugated polypeptides  
PT  
XX  
XX Claim 2; Page 44; 69pp; English.  
XX  
XX The invention provides immunogenic conjugated polypeptides which comprise  
CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory  
CC peptide and which promote binding to T cells. The novel immunogenic  
CC conjugated polypeptides are effective as an immunogen in a vaccine for  
CC treatment or prevention of infection by HSV and are represented by the  
CC formula PI-x-P2 or P2-x-PI, where PI = a HSV specific antigenic peptide  
CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein  
CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and  
CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of  
CC an immunoprotein which promotes binding to a class of subclass of T cells  
CC and which directs a predominantly TH1 type immune response to the peptide  
CC PI; and x = a covalent bond or a cleavable or non-cleavable peptide  
CC linking group. The conjugated polypeptides can elicit an immune response  
CC for neutralizing HSV and killing HSV infected cells. They can be used for

CC the treatment or prevention of HSV infection. In addition, DNA encoding  
CC the polypeptide can be used for genetic immunisation. They can also be  
CC used to diagnose the presence of infection, active or latent, in an  
CC individual by HSV by mixing T cells from the individual with the  
CC polypeptide and detecting a reaction between the T cells and the  
CC polypeptide.  
XX  
SQ Sequence 8 AA;  
Query Match 100.0%; Score 36; DB 20; Length 8;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SSIERFARL 8  
DB 1 SSIERFARL 8  
RESULT 3  
AAV10350  
ID AAV10350 standard; peptide; 8 AA.  
AC AAV10350;  
XX 12-MAY-1999 (first entry)  
DT  
DE T cell epitope/MHC ligand SEQ ID NO:280.  
XX  
XX Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;  
KW immunisation; tumour; infectious disease; immunotherapy; cancer;  
KW malignant melanoma; viral disease; hepatitis; AIDS.  
XX  
OS Synthetic.  
OS Herpes simplex virus.  
XX  
XX WO99022183-A2.  
XX  
XX 21-JAN-1999.  
PD  
XX 10-JUL-1998; 98WO-US14289.  
PF  
XX 10-DEC-1997; 97US-0988320.  
PR  
XX 10-JUL-1997; 97CA-2209815.  
XX  
XX (CTLI-) CTL IMMUNOTHERAPIES CORP.  
XX  
XX Kuendig TM, Simard JTL;  
PI  
XX WPI; 1999-120514/10.  
DR  
XX  
XX Inducing a cytotoxic T lymphocyte response - by maintaining a level  
PT of antigen in the lymphatic system of a mammal so as to provide a  
PT sustained CTL response, used to treat, e.g. AIDS  
XX  
XX Disclosure; Page 35; 1999p; English.  
XX  
XX The present invention describes a method of inducing and/or sustaining  
CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The  
CC method comprises: (a) delivering an antigen to the mammal at a level to  
CC induce an immunological CTL response in the mammal; and (b) maintaining  
CC the level of the antigen in the mammal's lymphatic system to maintain  
CC the immunologic CTL response. The method can be used for the delivery of  
CC e.g. a differentiation antigen, a tumour-specific multineage antigen,  
CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor  
CC gene antigen, or a viral antigen. They can be used for the treatment of  
CC disease such as cancer, e.g. malignant melanoma or infectious disease,  
CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery  
CC to the lymphatic system provides for potent CTL stimulation that takes  
CC place in the milieu of the lymphoid organ, and it sustains stimulation  
CC that is necessary to keep CTL active, cytotoxic and recirculating  
CC through the body. AAV10071 to AAV10639 represent examples of peptide  
CC antigens given in the present invention.  
XX

```
SO Sequence      8 AA;          100.0%; Score 36; DB 20; Length 8;
Query Match     Best Local Similarity    100.0%; Pred. No.:7.8e+05;
Matches         8; Conservative       0; Mismatches   0; Indels   0; Gaps   0;
```

```
OY           1 SSIPEARL 8
              |||||||
Db            1 SSIEPARL 8
```

```
RESULT 4
AAB12542 ID ABB12542 standard; Peptide: 8 AA.
```

```
AAC AAB12542;
ADT 03-NOV-2000 (first entry)
DEB gb-specific H-2b restricted peptide SEQ ID NO:2.
DSE Herpes virus; HSV; antibacterial; antiviral; antiparasitic; neurotropic;
KMW anti-diabetic; neuropsychiatric; antiparkinsonian; vaccine; gene therapy;
KKM amplicon; packaging; delivery.
XX Synthetic.
OS WO200034497-A2.
PN 15-JUN-2000.
PD PF 09-DEC-1999; 99WO-US29120.
PR XX 09-DEC-1998; 98US-O111630.
XX PR 26-Oct-1999; 99US-O161374.
XA PA (BREX/) BREAKEFIELD X O.
PA PA (CHIO/) CHIOCCE E A.
PA PA (SAEK/) SAEEKI Y.
PA PA (FRAEF/) FRAEPEL C.
PA PA (TOBL/) TOBLER K.
PA PA (ACKR/) ACKERMANN M.
PA PA (SUTR/) SUTTER M.
PA PA (ADEM/) ADEMA G J.
PA PA (SHOR/) SHORTMAN K.
```

```
P1 Breakefield XO, Chiocca EA, Saeki Y, Fraefel C, Tobler K;
P1 Ackermann M, Suter M, Adema GJ, Shortman K;
DR WP1: 2000-431309/37.
```

```
PT Herpes virus amplicon packaging system comprising herpes virus amplicon
PT vector and packaging vector having single cloning vehicle containing
PT replication proficient but packaging defective entire herpes virus
PT genome -
PS Example 3; Page 49; 78pp; English.
```

The present invention describes a Herpes virus (HSV) amplicon packaging system (I) comprising a HSV amplicon vector (II) and a packaging vector (III) comprising a single cloning vehicle which contains a replication deficient and packaging defective entire HSV genome. (II) is packaged into infectious particles that are substantially (III) of helper virus contamination by co-transfection with (III) which provides helper virus function. (I) is used for generating recombinant HSV vectors which are used for inducing protective immunity against herpes simplex virus (HSV-1) and any viral, bacterial or parasitic pathogen in an animal. The defective HSV-1 genome packaging system can be used to package a wide range of desired nucleotide segments, preferably a DNA segment, into an empty HSV particle. The HSV particles can be used to deliver heterologous DNA to a target cell. The heterologous sequence can encode any desired protein such as a therapeutic protein e.g. one that compensates for an inherited or acquired deficiency, e.g. tyrosine

CC hydroxylase for the treatment of Parkinson's disease, neurotrophic  
CC factors including neurotrophins, for the treatment of Alzheimer's  
CC disease, nerve growth factor receptor and the Trk receptor, hypoxanthine-  
CC guanine phosphoribosyl transferase (HGPRT) for the treatment of Lesch  
CC Nyhan disease, beta-hexosaminidase alpha chain for the treatment of Tay  
CC Sachs disease, insulin for the treatment of diabetes. The present  
CC sequence represents a peptide used in an example from the present  
CC invention.  
XX XX

Sequence 8 AA;  
SQ

Query Match 100.0%; Score 36; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8  
|||||||  
Db 1 SSIFARL 8

RESULT 5  
AA67358  
ID AAY67358 standard; peptide; 8 AA.

XX AAY67358;  
AC  
XX  
XX 25-APR-2000 (first entry)  
DT  
XX HSV glycoprotein B, SSTI peptide sequence used as a target peptide.  
DE

XX Target peptide; cytotoxic T lymphocyte; CTL; CTL immune response;  
KW cellular immune response induction method; vaccine; human; tumour;  
MW glycoprotein B.  
XX  
XX Herpes simplex virus.  
OS  
XX  
XX MO9963945-A2.  
PN  
PD 16-DEC-1999.  
XX  
XX 11-JUN-1999; 99WO-US13146.  
PF  
XX 12-JUN-1998; 98US-0089055.  
PR 30-OCT-1998; 98DS-0106339.  
XX  
XX (SLOK ) SIOAN KETTERING INST CANCER RES.  
PA  
XX Nikolic-Zugic J, Dyall R, Houghton AN;  
PI WPI; 2000-126432/11.  
DR  
XX  
XX

Induction of a cellular immune response to a weakly immunogenic  
protein, used to target and kill tumour cells -  
PT  
XX  
PS Claim 16; Page 24; 44pp; English.

This sequence represents a Herpes simplex virus glycoprotein B peptide  
used as a target peptide in the method of the invention. The invention  
relates to a method for inducing a cytotoxic T lymphocyte (CTL) immune  
response to non/weakly-immunogenic proteins which are expressed on tumour  
cells. The method for inducing a cellular immune response to a  
non-immunogenic or weakly immunogenic target peptide expressed on tumour  
cells of a mammalian subject comprises administering antigen to induce a  
cellular immune response to the target peptide. The antigen comprises an  
immunogenic portion having a major histocompatibility complex (MHC)  
binding domain which binds to the MHC and an immune recognition domain  
which is recognized by T-cells. The antigen is derived from the target  
peptide such that the MHC-binding portion binds to MHC with a greater  
affinity than the target peptide without material alteration of the  
immune recognition portion. The methods are used for inducing a cellular  
immune response to a non-immunogenic or weakly immunogenic target peptide  
expressed on tumour cells of a mammalian subject. The antigens and  
immunogens of the invention, as well as polynucleotides encoding them,

CC are used in vaccine compositions against tumour cells.  
XX  
SQ Sequence 8 AA:

Query Match 100.0%; Score 36; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
DB 1 SSIEFARL 8

RESULT 6  
AAE13452  
ID AAE13452 standard; peptide: 19 AA.

AC AAE13452;  
DE 12-FEB-2002 (first entry)

XX Herpes simplex virus MHC class I peptide antigen #4.  
XX  
XX Antigenic complex; epitope: heat shock protein; HSP; tether; javelin;  
KW major histocompatibility complex; MHC; therapy; immune response;  
KW malignancy.

OS Herpes simplex virus.

XX Key Location/Qualifiers

FT Region 1..8  
FT Region /note= "Javelin sequence"

FT Region 9..11  
FT Region /note= "Linker"

FT Region 12..19  
FT Region /note= "MHC class I epitope"

PN MO200179259-A1.

PD 25-OCT-2001.

PF 17-APR-2001; 2001WO-US12567.

PR 17-APR-2000; 2000US-197462P.

PA (ROTH/) ROTHMAN J E.  
PA (MAYH/) MAYHEW M.

PA (HOEM/) HOE M.

PI Rothman JE, Mayhew M, Hoe M;

DR WPI: 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a  
XX heat shock protein by a molecular tether designated a javelin are  
XX useful to treat or prevent infectious disease or malignancy -  
XX  
XX Disclosure: Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a  
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by  
XX a tethering molecule referred to as javelin which has affinity for the  
XX HSP under physiological conditions, where the epitopes are covalently  
XX joined to the tethering molecule and one epitope is major  
XX histocompatibility complex class I (MHC) and the other MHC class II. The  
XX antigenic complex is used to induce immune responses directed towards the  
XX treatment or prevention of infectious diseases and malignancies. The  
XX present sequence is Herpes simplex virus MHC class I peptide antigen.

SQ Sequence 19 AA:

Query Match 100.0%; Score 36; DB 23; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
DB 12 SSIEFARL 19

RESULT 7  
AAE13453  
ID AAE13453 standard; peptide: 19 AA.

AC AAE13453;  
DE 12-FEB-2002 (first entry)

XX Herpes simplex virus MHC class I peptide antigen #5.  
XX  
XX Antigenic complex; epitope: heat shock protein; HSP; tether; javelin;  
KW major histocompatibility complex; MHC; therapy; immune response;  
KW malignancy.

XX Herpes simplex virus.  
XX  
XX Key Location/Qualifiers  
FT Region 1..8  
FT Region /note= "MHC class I epitope"

FT Region 9..11  
FT Region /note= "Linker"

FT Region 12..19  
FT Region /note= "Javelin sequence"

PN MO200179259-A1.

PD 25-OCT-2001.

PF 17-APR-2001; 2001WO-US12567.

PR 17-APR-2000; 2000US-197462P.

PA (ROTH/) ROTHMAN J E.  
PA (MAYH/) MAYHEW M.

PA (HOEM/) HOE M.

PI Rothman JE, Mayhew M, Hoe M;

DR WPI: 2002-017594/02.

XX A new antigenic complex comprising epitopes non-covalently joined to a  
XX heat shock protein by a molecular tether designated a javelin are  
XX useful to treat or prevent infectious disease or malignancy -  
XX  
XX Disclosure: Page 14; 47pp; English.

XX The present invention relates to an antigenic complex, comprising a  
XX number of epitopes non-covalently joined to a heat shock protein (HSP) by  
XX a tethering molecule referred to as javelin which has affinity for the  
XX HSP under physiological conditions, where the epitopes are covalently  
XX joined to the tethering molecule and one epitope is major  
XX histocompatibility complex class I (MHC) and the other MHC class II. The  
XX antigenic complex is used to induce immune responses directed towards the  
XX treatment or prevention of infectious diseases and malignancies. The  
XX present sequence is Herpes simplex virus MHC class I peptide antigen.

SQ Sequence 19 AA:

Query Match 100.0%; Score 36; DB 23; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
DB 1 SSIEFARL 8

**RESULT 8**

ID AAU82071 standard; peptide; 22 AA.

AC AAU82071;

DT 09-APR-2002 (first entry)

DE Antigenic peptide extgB1 associated with Herpes Simplex Virus

KM T-cell binding ligand; TCBL; peptide G'; humanCMC class II beta chain;  
KM  
KM  
KM peptidic J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;  
KM immunological disorder; immune response; human immunodeficiency virus;  
KM herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA  
KM acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;  
KM autoimmune myocarditis; cytostatic; antineoplastic.

OS Synthetic

PN W0200189286-A2

PD 29-NOV-2001.

PF 24-MAY-2001; 2001WO-US16793.

PR 24-MAY-2000; 2000US-206548P.

PA (CELS-) CEL-SCI CORP.

PI Zimmerman DS, Sarin PS;

DR WPI; 2002-083037/11.

PT New T cell binding ligand peptide for treating immunological disorders  
PT such as herpes simplex virus, tuberculosis, cancers, acquired  
PT immunodeficiency syndrome and allergies

PS Disclosure; Page 27; 110pp; Engl1sh.

The present invention relates to novel T-cell binding ligand (TCBL) peptides (e.g. peptide G' (modified human MHC class II beta chain peptide G, peptide J (human beta-2-microglobulin peptide), and HIV-1 peptides). The TCBL peptide constructs are useful for treating immunological disorders. The peptide constructs are useful for eliciting a cellular immune response in a human patient. The method comprises administering the peptide construct to the patient preferably in combination with an immune response adjuvant. The peptide constructs in the form of conjugated peptides are useful for eliciting a cellular immune response in a patient exposed to or at risk for exposure to the human immunodeficiency virus (HIV). The TCBL peptides are useful for treating a patient suffering from an immunological disorder such as herpes simplex virus (HSV) infection, malaria, tuberculosis, cancers, acquired immunodeficiency syndrome (AIDS), allergies, autoimmune diseases (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune myocarditis, diabetes and lupus) by administering a peptide construct comprising a TCBL peptide bonded to an antigenic peptide associated with the disorder. Unlike prior art peptide conjugates, a modified version of peptide G has long range stabilisation and also enhances the immune response. AA082019-AA082014 represent T-cell specific binding ligand peptides, peptide constructs or peptides used in their construction.

**SQ Sequence 22 AA;**

Query Match	100.0%	Score 36;	DB 23;	Length 22;
Best Local Similarity	100.0%	Pred. NO.	0.12;	
Matches	8;	Conservative	0;	Indels 0;
		Mismatches	0;	Gaps 0;

QY	1	SSIEFARL	8
Db	7	SSIEFARL	14

RESULT 9  
PAY12477

ID	AA	13421	standard; peptide; 23 AA.
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

AC MAY13421;

DT 19-JUL-1999 (first entry)

Conjugated control peptide L-B1

KM Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICR4  
 KW Immunomodulatory; T cell; Immunogen; vaccine; ICP27; glycoprotein B;  
 KW ribonucleotide reductase; ICR3a.5; glycoprotein E; glycoprotein F;  
 KW Immune response; genetic immunisation:

OS Synthetic.

PN WO9916710-A1

PD 08-APR-1999

PF 29-SEP-1998; 98WO-US20681.

PR 30-SEP-1997; 97US-0060422.

PA (CELS-) CEL-SCI CORP.  
PA (UYNE-) UNIV NORTHEASTERN OHIO,  
PA

PI Rosenthal KS, Zimmerman DH;

DR WPT; 1999-312418/26.

**PT** New immunogenic conjugated polypeptides

PS Example 1; Page 33; 69pp; English.

The invention provides immunogenic conjugated polypeptides which comprise a herpes simplex virus peptide (HSV) linked to an immunomodulatory peptide and which promote binding to T cells. The novel immunogenic conjugated polypeptides are effective as an immunogen in a vaccine for treatment or prevention of infection by HSV and are represented by the formula p1-x-p2 or p2-x-p1: where p1 = a HSV specific antigenic peptide from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and glycoprotein F; p2 = an immunomodulatory peptide which is a portion of an immunoprotein which promotes binding to a class of subclass of T cells and which directs a predominantly TH1 type immune response to the peptide p1; and x = a covalent bond or a cleavable or non-cleavable peptide linking group. The conjugated polypeptides can elicit an immune response for neutralizing HSV and killing HSV infected cells. They can be used for the treatment or prevention of HSV infection. In addition, DNA encoding the polypeptide can be used for genetic immunization. They can also be used to diagnose the presence of infection, active or latent, in an individual by HSV by mixing T cells from the individual with the polypeptide and detecting a reaction between the T cells and the polypeptide.

50 Sequence 23 AA;

Query Match	100.0%	Pred. 36;	DB 20;	Length 23;
Best Local Similarity	100.0%	Pred. No. 0.13;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

QY	1	SSIEFARL	8
Db	16	SSIEFARL	23

RESULT	ID
AAV13420	standard; peptide; 26 AA
XX	
AC	AAV13420;

XX 19-JUL-1999 (first entry)  
 XX Conjugated peptide G-B1.  
 DE  
 XX Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;  
 KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;  
 KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;  
 KW immune response; genetic immunisation.  
 XX Synthetic.  
 XX OS  
 XX MO9916710-A1.  
 XX 08-APR-1999.  
 XX 29-SEP-1998; 98WO-US20681.  
 XX 30-SEP-1997; 97US-0060422.  
 XX (CELS-) CEL-SCI CORP.  
 XX (UYNE-) UNIV NORTHEASTERN OHIO.  
 XX Rosenthal KS, Zimmerman DH;  
 XX WPI; 1999-312418/26.  
 XX New Immunogenic conjugated polypeptides  
 PS Claim 4; Page 45; 69pp; English.  
 XX The invention provides immunogenic conjugated polypeptides which comprise  
 CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory  
 CC peptide and which promote binding to T cells. The novel immunogenic  
 CC conjugated polypeptides are effective as an immunogen in a vaccine for  
 CC treatment or prevention of infection by HSV and are represented by the  
 CC formula P1-x-P2 or P2-x-P1, where P1 = a HSV specific antigenic peptide  
 CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein  
 CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and  
 CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of  
 CC an immunoprotein which promotes binding to a class of subclass of T cells  
 CC and which directs a predominantly TH1 type immune response to the peptide  
 CC P1; and x = a covalent bond or a cleavable or non-cleavable peptide  
 CC linking group. The conjugated polypeptides can elicit an immune response  
 CC for neutralizing HSV and killing HSV infected cells. They can be used for  
 CC the treatment or prevention of HSV infection. In addition, DNA encoding  
 CC the polypeptide can be used for genetic immunization. They can also be  
 CC used to diagnose the presence of infection, active or latent, in an  
 CC individual by HSV by mixing T cells from the individual with the  
 CC polypeptide and detecting a reaction between the T cells and the  
 CC polypeptide.  
 CC  
 CC Sequence 26 AA;  
 XX  
 SO  
 Query Match 100.0%; Score 36; DB 20; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSIFFARL 8  
 Db 19 SSIFFARL 26  
 |||||  
 RESULT 11  
 AAY13419  
 ID AAY13419 standard; peptide; 27 AA.  
 XX  
 AC AAY13419;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE Conjugated peptide J-B1.  
 XX

KW Immunogenic; conjugated polypeptide; herpes simplex virus; HSV; ICP4;  
 KW immunomodulatory; T cell; immunogen; vaccine; ICP27; glycoprotein B;  
 KW ribonucleotide reductase; ICP34.5; glycoprotein E; glycoprotein F;  
 KW immune response; genetic immunisation.  
 XX Synthetic.  
 XX OS  
 XX Herpes simplex virus type 1.  
 XX MO9916710-A1.  
 XX 08-APR-1999.  
 XX 29-SEP-1998; 98WO-US20681.  
 XX 30-SEP-1997; 97US-0060422.  
 XX (CELS-) CEL-SCI CORP.  
 XX (UYNE-) UNIV NORTHEASTERN OHIO.  
 XX Rosenthal KS, Zimmerman DH;  
 XX WPI; 1999-312418/26.  
 XX New Immunogenic conjugated polypeptides  
 PS Claim 4; Page 45; 69pp; English.  
 XX The invention provides immunogenic conjugated polypeptides which comprise  
 CC a herpes simplex virus peptide (HSV) linked to an immunomodulatory  
 CC peptide and which promote binding to T cells. The novel immunogenic  
 CC conjugated polypeptides are effective as an immunogen in a vaccine for  
 CC treatment or prevention of infection by HSV and are represented by the  
 CC formula P1-x-P2 or P2-x-P1, where P1 = a HSV specific antigenic peptide  
 CC from a protein of HSV type 1 or type 2, selected from ICP27, glycoprotein  
 CC B, ribonucleotide reductase, ICP4, ICP34.5, glycoprotein E and  
 CC glycoprotein F; P2 = an immunomodulatory peptide which is a portion of  
 CC an immunoprotein which promotes binding to a class of subclass of T cells  
 CC and which directs a predominantly TH1 type immune response to the peptide  
 CC P1; and x = a covalent bond or a cleavable or non-cleavable peptide  
 CC linking group. The conjugated polypeptides can elicit an immune response  
 CC for neutralizing HSV and killing HSV infected cells. They can be used for  
 CC the treatment or prevention of HSV infection. In addition, DNA encoding  
 CC the polypeptide can be used for genetic immunization. They can also be  
 CC used to diagnose the presence of infection, active or latent, in an  
 CC individual by HSV by mixing T cells from the individual with the  
 CC polypeptide and detecting a reaction between the T cells and the  
 CC polypeptide.  
 CC  
 CC Sequence 27 AA;  
 XX  
 SO  
 Query Match 100.0%; Score 36; DB 20; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSIFFARL 8  
 Db 20 SSIFFARL 27  
 |||||  
 RESULT 12  
 AAE13454  
 ID AAE13454 standard; peptide; 30 AA.  
 XX  
 AC AAE13454;  
 XX  
 DT 12-FEB-2002 (first entry)  
 XX  
 DE Herpes simplex virus MHC class I peptide antigen #6.  
 XX  
 AC Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;  
 KW major histocompatibility complex; MHC; therapy; immune response;  
 KW malignancy.  
 XX

OS Herpes simplex virus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..8  
 FT /note- "Javelin sequence"  
 FT Region 9..11  
 FT /note- "Linker"  
 FT Region 12..19  
 FT /note- "MHC class I epitope"  
 FT Region 20..22  
 FT /note- "Linker"  
 FT Region 23..30  
 FT /note- "Javelin sequence"  
 XX  
 PN WO200179259-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 17-APR-2001; 2001WO-US12567.  
 XX  
 PR 17-APR-2000; 2000US-197462P.  
 XX  
 XX (ROTH/) ROTHMAN J E.  
 PA (MAYR/) MAYHEW M.  
 PA (HOEM/) HOE M.  
 XX  
 PI Rothman JE, Mayhew M, Hoe M;  
 XX  
 DR WPI; 2002-017594/02.  
 XX  
 PT A new antigenic complex comprising epitopes non-covalently joined to a  
 PT heat shock protein by a molecular tether designated a javelin are  
 PT useful to treat or prevent infectious disease or malignancy  
 XX  
 PS Disclosure; Page 14; 47pp; English.  
 XX  
 CC The present invention relates to an antigenic complex, comprising a  
 CC number of epitopes non-covalently joined to a heat shock protein (HSP) by  
 CC a tethering molecule referred to as javelin which has affinity for the  
 CC HSP under physiological conditions, where the epitopes are covalently  
 CC joined to the tethering molecule and one epitope is a major  
 CC histocompatibility complex class I (MHC) and the other MHC class II. The  
 CC antigenic complex is used to induce immune responses directed towards the  
 CC treatment or prevention of infectious diseases and malignancies. The  
 CC present sequence is Herpes simplex virus MHC class I peptide antigen.  
 CC  
 SQ Sequence 30 AA;  
 XX  
 Query Match 100.0%; Score 36; DB 23; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SSIEFARL 8  
 DB 12 SSIEFARL 19  
 XX  
 RESULT 13  
 AAU82093  
 ID AAU82093 standard; peptide; 38 AA.  
 XX  
 AC AAU82093;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE T-cell specific binding ligand peptide construct #27.  
 XX  
 KW T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;  
 KW peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;  
 KW immunological disorder; immune response; human immunodeficiency virus;  
 KW herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;  
 KW acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;  
 KW autoimmune myocarditis; cytostatic; antineoplastic; antineoplastic.

XX  
 OS Synthetic.  
 XX  
 PN WO200189286-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001WO-US16793.  
 XX  
 PR 24-MAY-2000; 2000US-206548P.  
 XX  
 PA (CELS-) CELS-SCI CORP.  
 XX  
 PI Zimmerman DS, Sarlin PS;  
 XX  
 DR WPI; 2002-083037/11.  
 XX  
 PD New T cell binding ligand peptide for treating immunological disorders  
 PT such as herpes simplex virus, tuberculosis, cancers, acquired  
 PT immunodeficiency syndrome and allergies  
 XX  
 PS Disclosure; Page 28; 110pp; English.  
 XX  
 CC The present invention relates to novel T-cell binding ligand (TCBL)  
 CC peptides (e.g. peptide G' (modified human MHC class II beta chain  
 CC peptide G, peptide J (human beta-2-microglobulin peptide) and HIV-1  
 CC peptides) and TCBL peptide constructs for treating immunological  
 CC disorders. The peptide constructs are useful for eliciting a cellular  
 CC immune response in a human patient. The method comprises administering  
 CC the peptide construct to the patient preferably in combination with an  
 CC immune response adjuvant. The peptide constructs in the form of  
 CC conjugated peptides are useful for eliciting a cellular immune response  
 CC in a patient exposed to or at risk for exposure to the human  
 CC immunodeficiency virus (HIV). The TCBL peptides are useful for treating a  
 CC patient suffering from an immunological disorder such as herpes simplex  
 CC virus (HSV) infection, malaria, tuberculosis, cancers, acquired  
 CC immunodeficiency syndrome (AIDS), allergies, autoimmune diseases  
 CC (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune  
 CC myocarditis, diabetes and lupus) by administering a peptide construct  
 CC comprising a TCBL peptide bonded to an antigenic peptide associated with  
 CC the disorder. Unlike prior art peptide conjugates, a modified version of  
 CC peptide G has long range stabilisation and also enhances the immune  
 CC response. AAU82019-AAU82114 represent T-cell specific binding ligand  
 CC peptides; peptide constructs or peptides used in their construction.  
 CC  
 SQ Sequence 38 AA;  
 XX  
 Query Match 100.0%; Score 36; DB 23; Length 38;  
 Best Local Similarity 100.0%; Pred. No. 0.22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SSIEFARL 8  
 DB 23 SSIEFARL 30  
 XX  
 RESULT 14  
 AAU82094  
 ID AAU82094 standard; peptide; 40 AA.  
 XX  
 AC AAU82094;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE T-cell specific binding ligand peptide construct #28.  
 XX  
 KW T-cell binding ligand; TCBL; peptide G'; human MHC class II beta chain;  
 KW peptide J; human beta-2-microglobulin; HIV-1; TCBL peptide construct;  
 KW immunological disorder; immune response; human immunodeficiency virus;  
 KW herpes simplex virus infection; HSV; malaria; tuberculosis; cancer; CEA;  
 KW acquired immunodeficiency syndrome; AIDS; allergy; autoimmune disease;  
 KW autoimmune myocarditis; cytostatic; antineoplastic; antineoplastic.

OS Synthetic.  
 XX  
 PN WO200189286-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PE 24-MAY-2001; 2001WO-US16793.  
 XX  
 PR 24-MAY-2000; 2000US-206548P.  
 XX  
 PA (CELS-) CEL-SCI CORP.  
 XX  
 PI Zimmerman DS, Sarin PS;  
 XX  
 DR WPI; 2002-083037/11.  
 XX  
 PT New T cell binding ligand peptide for treating immunological disorders  
 PT such as herpes simplex virus tuberculosis, cancers, acquired  
 PT immunodeficiency syndrome and allergies  
 XX  
 PS Disclosure; Page 28; 110pp; English.  
 XX  
 CC The present invention relates to novel T-cell binding ligand (TCBL)  
 CC peptides (e.g. peptide G' (modified human MHC class II beta chain  
 CC peptide G, peptide J (human beta-2-microglobulin peptide) and HIV-1  
 CC peptides) and TCBL peptide constructs are useful for eliciting a cellular  
 CC disorders. The peptide constructs are useful for eliciting a cellular  
 CC immune response in a human patient. The method comprises administering  
 CC the peptide construct to the patient preferably in combination with an  
 CC immune response adjuvant. The peptide constructs in the form of  
 CC conjugated peptides are useful for eliciting a cellular immune response  
 CC in a patient exposed to or at risk for exposure to the human  
 CC immunodeficiency virus (HIV). The TCBL peptides are useful for treating a  
 CC patient suffering from an immunological disorder such as herpes simplex  
 CC virus (HSV) infection, malaria, tuberculosis, cancers, acquired  
 CC immunodeficiency syndrome (AIDS), allergies, autoimmune diseases  
 CC (e.g. arthritis, Graves disease, multiple sclerosis (MS), autoimmune  
 CC myocarditis, diabetes and lupus) by administering a peptide construct  
 CC comprising a TCBL peptide bonded to an antigenic peptide associated with  
 CC the disorder. Unlike prior art peptide conjugates, a modified version of  
 CC peptide G has long range stabilisation and also enhances the immune  
 CC response. AAU82019-AAU82114 represent T-cell specific binding ligand  
 CC peptides, peptide constructs or peptides used in their construction.  
 XX  
 SQ Sequence 40 AA:  
 Query Match 100.0%; Score 36; DB 23; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 0.24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSIFFARL 8  
 Db 25 SSIFFARL 32  
 RESULT 15  
 AARI4666  
 ID AARI4666 standard; Protein; 694 AA.  
 XX  
 AC AARI4666;  
 XX  
 DT 23-JAN-1992 (first entry)  
 XX  
 DE Truncated HSVgB polypeptide.  
 XX  
 KW Vaccine; antigen.  
 XX  
 OS Herpes simplex virus.  
 XX  
 PN JP03218397-A.  
 XX  
 PD 25-SEP-1991.  
 XX

PF 21-JUN-1990; 90JP-0161448.  
 XX  
 PR 30-NOV-1989; 89JP-0308941.  
 PR 22-JUN-1989; 89JP-0158238.  
 PR 21-JUN-1990; 90JP-0161448.  
 XX  
 PA (TAKE ) TAKEDA CHEMICAL IND KK.  
 XX  
 DR WPI; 1991-328397/45.  
 DR N-PSDB; AAQ14479.  
 XX  
 PT HSVgB polypeptide(s) obtd. by recombinant DNA techniques -  
 PT useful as vaccines against HSV and in diagnosis, can be produced  
 PT cheaply and safely.  
 XX  
 PS Claim 1; Fig 7; 24pp; Japanese.  
 XX  
 CC The sequence was deduced from DNA and is that of a truncated HSVgB  
 CC polypeptide. The recombinant protein can be used to prepare  
 CC vaccines for prophylaxis of HSV infection and for use in diagnostic  
 CC kits.  
 CC See also AARI4665.  
 XX  
 SQ Sequence 694 AA:  
 Query Match 100.0%; Score 36; DB 12; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SSIFFARL 8  
 Db 470 SSIFFARL 477

Search completed: May 22, 2003, 12:08:18  
 Job time : 33 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:06:12 ; Search time 14 Seconds

(without alignments)  
54.934 Million cell updates/sec

Title: US-09-719-494-10

Perfect score: 36

Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*\*\*  
2: PIR1:\*\*\*  
3: PIR2:\*\*\*  
4: PIR3:\*\*\*  
5: PIR4:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	903	1 VGBEB1	glycoprotein B pre
2	36	100.0	903	1 VGBEB1	glycoprotein B pre
3	36	100.0	904	1 VGBEB7	glycoprotein B pre
4	36	100.0	904	1 VGBEB2	glycoprotein B pre
5	36	100.0	904	1 VGBEB2	glycoprotein B pre
6	35	97.2	885	1 VGBESA	glycoprotein B pre
7	35	97.2	917	1 VGBEBH	glycoprotein B pre
8	31	86.1	920	1 D48349	glycoprotein B pre
9	30	83.3	83	2 T36350	probable regulator
10	30	83.3	153	2 F81429	hypothetical prote
11	30	83.3	344	1 S46020	probable aldehyde
12	30	83.3	350	2 T21106	hypothetical prote
13	30	83.3	465	1 F64630	cysteine-tRNA ligase
14	30	83.3	465	1 D71884	cysteine-tRNA ligase
15	30	83.3	919	1 VGBE0H	glycoprotein B pre
16	30	83.3	975	2 T42576	probable envelope
17	30	83.3	979	2 JH0109	glycoprotein 14 pr
18	30	83.3	980	1 VGBE2H	glycoprotein B pre
19	30	83.3	980	1 VGBE2H	glycoprotein B pre
20	30	83.3	1315	2 T50262	probable nucleopor
21	29	80.6	154	2 S03242	hypothetical prote
22	29	80.6	208	2 S34338	leukotoxin A - Pas
23	29	80.6	288	1 BYECRA	repa protein - Esc
24	29	80.6	317	2 AF0582	probable permease
25	29	80.6	318	2 C70636	probable echal3 pr
26	29	80.6	321	2 T26974	hypothetical prote
27	29	80.6	452	2 AB3222	two component sens
28	29	80.6	532	2 H84123	CTP synthetase ctr
29	29	80.6	608	2 H81706	oligodeopeptidase

#### ALIGNMENTS

```

RESULT 1
VGBEB1
glycoprotein B precursor - human herpesvirus 1 (strain F)
C:Species: human herpesvirus 1
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03750
R:Pelletier, P.E.; Kousoulas, K.G.; Pereira, L.; Roizman, B.
J. Virol. 53, 243-253, 1985
A:Title: Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene: primary
A:Reference number: A03750; MIMD:85083254; PMID:2981343
A:Accession: A03750
A:Molecule type: DNA
A:Residues: 1-903 <PELT>
A:Cross-references: GB:M14164; GB:M12398; NID:g330084; PIDN:AAA5776.1; PID:g330086
C:Superfamily: herpesvirus glycoprotein B
C:Keywords: glycoprotein; transmembrane protein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-903/Product: glycoprotein B #status predicted <TM>
F:726-746/Domain: transmembrane #status predicted <TM>
F:751-771/Domain: transmembrane #status predicted <TM>
F:774-794/Domain: transmembrane #status predicted <TM>
F:86,140,254,397,429,477,488,673,818,887/Binding site: carbohydrate (asn) (covalent
F:115-572,132-528,206-270,363-411,595-653/Disulfide Bonds: #status predicted
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Query Match 100.0%; Score 36; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 2.3;
Db 498 SSIEFARL 505
OY 1 SSIEFARL 8
IIIIIIII
Db 498 SSIEFARL 505
RESULT 2
VGBEB1
glycoprotein B precursor - human herpesvirus 1 (strain KOS)
C:Species: human herpesvirus 1
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03751
R:Beitz, D.J.; Fox, B.A.; Deluca, N.A.; Person, S.
Virology 133, 301-314, 1984
A:Title: Nucleotide sequence specifying the glycoprotein gene, gb, of herpes simple
A:Reference number: A03751; MIMD:84174058; PMID:6324454
A:Accession: A03751
A:Molecule type: DNA
A:Residues: 1-903 <BET>
A:Cross-references: GB:K01760; NID:g330082; PIDN:AAA5774.1; PID:g330083
C:Superfamily: herpesvirus glycoprotein B
C:Keywords: glycoprotein; transmembrane protein
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-903/Product: glycoprotein B #status predicted <TM>
F:726-746/Domain: transmembrane #status predicted <TM>

```

F:751-771/Domain: transmembrane #status predicted <TM2>  
 F:774-794/Domain: transmembrane #status predicted <TM3>  
 F:76, 86, 140, 256, 397, 429, 477, 488, 673, 818, 887/Binding site: carbohydrate (Asn) (covalent) #status  
 F:115-572, 132-528, 206-270, 353-411, 595-632/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 903;  
 Best Local Similarity 100.0%; Pred. No. 2,3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SSIFFARL 8  
 Db 498 SSIFFARL 505

## RESULT 3

VGEBE7  
 glycoprotein B precursor - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1  
 C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000

C:Accession: J13084  
 R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perr  
 J. Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes sin  
 A:Reference number: A50853; MUID:86274327; PMID:2839594

A:Accession: J13084  
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-904 <MCG>  
 A:Cross-references: GB:X14112; GB:D00317; GB:D00374; GB:S40593; NID:G1944556; PIDN:CAA32

C:Genetics:

A:Gene: UL27  
 C:Superfamily: herpesvirus glycoprotein B

C:Keywords: glycoprotein; transmembrane protein  
 F:1-30/Domain: signal sequence #status predicted <SIG>

F:31-904/Product: glycoprotein B #status predicted <MAT>  
 F:727-746/Domain: transmembrane #status predicted <TM1>

F:752-771/Domain: transmembrane #status predicted <TM2>  
 F:774-795/Domain: transmembrane #status predicted <TM3>

F:87, 141, 256, 398, 430, 478, 489, 674, 819, 888/Binding site: carbohydrate (Asn) (covalent) #st  
 F:116-573, 133-529, 207-271, 364-412, 596-633/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 2,3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SSIFFARL 8  
 Db 499 SSIFFARL 506

## RESULT 4

VGEBE2  
 glycoprotein B precursor - human herpesvirus 2 (strain HG52)

C:Species: human herpesvirus 2  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25611  
 R:Bzik, D.J.; Debroy, C.; Fox, B.A.; Pederson, N.E.; Person, S.

A:Title: The nucleotide sequence of the gb glycoprotein gene of HSV-2 and comparison wit  
 A:Reference number: A25611; MUID:87071654; PMID:3024391

A:Accession: A25611  
 A:Molecule type: DNA

A:Residues: 1-904 <BZI>  
 A:Cross-references: GB:M14923; NID:G330254; PIDN:AAA66440.1; PID:G330255

R:Norais, N.; Tang, D.; Kau, S.; Chamberlain, S.H.; Mastarz, F.R.; Burke, R.L.; Marcus,  
 J. Virol. 70, 7379-7387, 1996

A:Title: Disulfide bonds of herpes simplex virus type 2 glycoprotein gb.  
 A:Reference number: A58366; MUID:97048015; PMID:8892856

A:Contents: annotation; tryptic peptide disulfide bond assignments  
 C:Superfamily: herpesvirus glycoprotein B

C:Keywords: glycoprotein; transmembrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-904/Product: glycoprotein B #status predicted <MAT>

F:724-744/Domain: transmembrane #status predicted <TM1>  
 F:749-769/Domain: transmembrane #status predicted <TM2>  
 F:777-792/Domain: transmembrane #status predicted <TM3>  
 F:82, 136, 250, 393, 425, 473, 486, 671/Binding site: carbohydrate (Asn) (covalent) #status  
 F:111-570, 128-526, 202-266, 359-407, 593-630/Disulfide bonds: #status experimental

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 2,3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SSIFFARL 8  
 Db 496 SSIFFARL 503

## RESULT 5

VGEBE2  
 glycoprotein B precursor - human herpesvirus 2 (strain 333)

C:Species: human herpesvirus 2  
 C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jul-1999

C:Accession: A26790; A45928  
 R:Stuve, L.L.; Brown-Shimer, S.; Pachl, C.; Najarian, R.; Dina, D.; Burke, R.L.

J. Virol. 61, 326-335, 1987  
 A:Title: Structure and expression of the herpes simplex virus type 2 glycoprotein gb  
 A:Reference number: A26790; MUID:87112925; PMID:3027364

A:Accession: A26790  
 A:Molecule type: DNA

A:Residues: 1-904 <STU>  
 A:Cross-references: GB:M15118; NID:G330256; PIDN:AAA45837.1; PID:G330257

R:Zwaagstra, J.C.; Leung, W.C.  
 Can. J. Microbiol. 33, 879-887, 1987

A:Title: The nucleotide sequence of herpes simplex virus type 2 (333) glycoprotein gi  
 A:Reference number: A45928; MUID:88079667; PMID:2446730

A:Accession: A45928  
 A:Molecule type: DNA

A:Residues: 1-94, 'AMP', 'T', '309-481', 'R', '483-609', 'W', '611-664', 'R', '666-904 <ZWA>  
 A:Cross-references: GB:M47711; NID:G341245; PIDN:AAA60540.1; PID:G623400

A:Note: the authors translated the codon ATG for residue 610 as Ile  
 C:Superfamily: herpesvirus glycoprotein B

C:Keywords: glycoprotein; transmembrane protein  
 F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-904/Product: glycoprotein B #status predicted <MAT>  
 F:724-744/Domain: transmembrane #status predicted <TM1>

F:749-769/Domain: transmembrane #status predicted <TM2>  
 F:777-792/Domain: transmembrane #status predicted <TM3>

F:82, 136, 250, 393, 425, 473, 486, 671/Binding site: carbohydrate (Asn) (covalent) #status  
 F:111-570, 128-526, 202-266, 359-407, 593-630/Disulfide bonds: #status predicted

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 2,3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SSIFFARL 8  
 Db 496 SSIFFARL 503

## RESULT 6

VGEBE2  
 glycoprotein B precursor - simian herpesvirus SA8 (strain B264)

C:Species: simian herpesvirus SA8  
 C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 16-Jul-1999

C:Accession: J01332  
 R:Botchers, K.; Melgert, W.; Buhk, H.J.; Ludwig, H.; Mankertz, J.

J. Gen. Virol. 72, 2299-2304, 1991  
 A:Title: Conserved domains of glycoprotein B (gb) of the monkey virus, simian agent 1

A:Reference number: J01332; MUID:91374035; PMID:1895066  
 A:Accession: J01332

A:Molecule type: DNA  
 A:Residues: 1-885 <BOR>

A:Cross-references: EMBL:X56935; NID:G60438; PIDN:CAA40256.1; PID:G60439  
 C:Superfamily: herpesvirus glycoprotein B  
 C:Keywords: glycoprotein; transmembrane protein

F;1-28/Domain: signal sequence #status predicted <SIG>  
 F;29-885/Product: glycoprotein B #status predicted <VMT>  
 F;112-732/Domain: transmembrane #status predicted <TM1>  
 F;137-752/Domain: transmembrane #status predicted <TM2>  
 F;760-780/Domain: transmembrane #status predicted <TM3>  
 F;68,122,379,411,655/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F;97-558,114-514,188-252,345-393,581-618/Disulfide bonds: #status predicted

Query Match 97.2% Score 35; DB 1; Length 885;  
 Best Local Similarity 87.5%; Pred. No. 3.9;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 484 SSVFARL 491

RESULT 7  
 VGBEBH  
 glycoprotein B precursor - bovine herpesvirus 2 (strain BMV)  
 C/Species: bovine herpesvirus 2  
 C/Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999  
 C/Accession: C29242  
 R/Hammerschmidt, W.; Contraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.  
 Virology 165, 388-405, 1988  
 A/Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus 2  
 A/Reference number: A94381, MUID:88306231; PMID:2841793  
 A/Accession: C29242  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-917 <HAM>  
 A/Cross-references: GB:M21628; NID:g330752; PIDN:AAA6053.1; PID:g330753  
 C/Superfamily: herpesvirus glycoprotein B  
 C/Keywords: glycoprotein; transmembrane protein  
 F;1-22/Domain: signal sequence #status predicted <SIG>  
 F;23-917/Product: glycoprotein B #status predicted <GPB>  
 F;578-594/Domain: transmembrane #status predicted <TM1>  
 F;770-786/Domain: transmembrane #status predicted <TM2>  
 F;795-811/Domain: transmembrane #status predicted <TM3>  
 F;48,110,164,278,421,453,505,564,692/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 97.2% Score 35; DB 1; Length 917;  
 Best Local Similarity 87.5%; Pred. No. 4;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 517 SSVFARL 524

RESULT 8  
 D48349  
 glycoprotein B precursor - salmuthine herpesvirus 1 (strain MV-5-4-PSL)  
 C/Species: salmuthine herpesvirus 1  
 C/Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Sep-1994  
 C/Accession: D48349  
 R/Ederle, R.; Black, D.  
 Arch. Virol. 129, 167-182, 1993  
 A/Title: Sequence analysis of herpes simplex virus gB gene homologs of two platyrrhine primates  
 A/Reference number: A48349; MUID:9328440; PMID:8385913  
 A/Accession: D48349  
 A/Molecule type: DNA  
 A/Residues: 1-920 <EBE>  
 A/Note: sequence extracted from NCBI backbone (NCBIN:129066, NCBIPI:129068)  
 C/Superfamily: herpesvirus glycoprotein B  
 C/Keywords: glycoprotein; transmembrane protein  
 F;1-35/Domain: signal sequence #status predicted <SIG>  
 F;36-920/Product: glycoprotein B #status predicted <VMT>  
 F;576-592/Region: hydrophobic  
 F;761-780/Region: transmembrane #status predicted <TM1>  
 F;787-804/Domain: transmembrane #status predicted <TM2>  
 F;98,119,152,409,441,683,733/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.1% Score 31; DB 1; Length 920;  
 Best Local Similarity 85.7%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 SSIEFARL 8  
 DB 509 SSVFARL 515

RESULT 9  
 T36350  
 probable regulator - Streptomyces coelicolor  
 C/Species: Streptomyces coelicolor  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C/Accession: T36350  
 R/Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, May 1999  
 A/Reference number: Z21575  
 A/Accession: T36350  
 A/Status: preliminary; translated from GB/EMBL/DDBB  
 A/Molecule type: DNA  
 A/Residues: 1-83 <COLI>  
 A/Cross-references: EMBL:AL049841; PIDN:CA842777.1; GSPDB:GN00070; SCOEDB:SCE9.31c  
 A/Experimental source: strain A3(2)  
 C/Genetics:  
 A/Gene: SCOEDB:SCE9.31c

Query Match 83.3% Score 30; DB 2; Length 83;  
 Best Local Similarity 75.0%; Pred. No. 4.7;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 33 SSVFARL 40

RESULT 10  
 F81429  
 hypothetical protein Cj0126c [imported] - Campylobacter jejuni (strain NCTC 11168)  
 C/Species: Campylobacter jejuni  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
 C/Accession: F81429  
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Ch.  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; B.  
 Nature 403, 665-668, 2000  
 A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveal.  
 A/Reference number: A81250; MUID:20150912; PMID:10688204  
 A/Accession: F81429  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-153 <PAR>  
 A/Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72610.1; PID:g6.  
 A/Experimental source: serotype O2, strain NCTC 11168  
 C/Genetics:  
 C/Superfamily: conserved hypothetical protein HI0033

Query Match 83.3% Score 30; DB 2; Length 153;  
 Best Local Similarity 75.0%; Pred. No. 9.3;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 83 TSIEFARL 90

RESULT 11  
 S46020  
 probable aldehyde reductase (EC 1.1.1.21) - Yeast (Saccharomyces cerevisiae)  
 N/Alternate names: hypothetical protein YBR1127; hypothetical protein YBR149w  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Apr-2002  
 C/Accession: S46020; S28668

R. Entian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.; Li, Z.; Niegemann, E.; Sché S.; Grunbein, R.; Hedges, D.; Klesau, P.; Korol, S.; Krebs, B.; Proft, M.; Slegers, K. submitted to the Protein Sequence Database, August 1994

A:Reference number: S46013

A:Accession: S46020

A:Molecule type: DNA

A:Residues: 1-344 <ENT>

A:Cross-references: EMBL:Z36018; NID:q536473; PIDN:CAA8107.1; PID:q536474; GSPDB:GN0000

A:Experimental source: strain S288C

R.Mattinez-Soriano, J.P.; Wong, W.M.; van Ryk, D.I.; Nazar, R.N.

J. Mol. Biol. 217, 629-635, 1991

A:Title: A widely distributed "cat" family of repetitive DNA sequences.

A:Reference number: S28668; MUID:91171289; PMID:2005616

A:Accession: S28668

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-55, 'LQSKLDN', 63-70, 'SR', 74-344 <MAR>

A:Cross-references: EMBL:M95580; NID:q172583; PIDN:AAA5037.1; PID:q172584

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992

C:Genetics:

A:Gene: SGD:ARAL; MIPS:YBR149V

A:Cross-references: SGD:S0000353

A:Map position: 2R

C:Superfamily: aldehyde reductase

C:Keywords: oxidoreductase

Query Match 83.3%; Score 30; DB 1; Length 344;  
Best Local Similarity 87.5%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSIERFARL 8  
|||||  
DB 294 SSIERFARL 301

RESULT 12

T21106

hypothetical protein F19B6.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T21106

R.Thomas, K.

submitted to the EMBL Data Library, February 1996

A:Reference number: 219375

A:Accession: T21106

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-350 <MIL>

A:Cross-references: EMBL:Z69635; NID:q1200023; PIDN:CAA93458.1; GSPDB:GN00022; CESP:F19B

A:Experimental source: clone F19B6

C:Genetics:

A:Gene: CESP:F19B6.3

A:Map position: 4

A:Introns: 180/3; 274/2

Query Match 83.3%; Score 30; DB 2; Length 350;  
Best Local Similarity 62.5%; Pred. No. 23;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIERFARL 8  
|||||  
DB 125 SSIERFARL 132

RESULT 13

F64630

cysteine-tRNA ligase (EC 6.1.1.16) - *Helicobacter pylori* (strain 26695)

N:Alternate names: cysteinyl-tRNA synthetase

C:Species: *Helicobacter pylori*

C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 03-Jun-2002

C:Accession: F64630; C53739

R.Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenna

son, J.D.; Kelley, J.N.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: F64630

A:Molecule type: DNA

A:Status: nucleic acid sequence not shown; translation not shown

A:Residues: 1-465 <TOM>

A:Cross-references: GB:AE00511; NID:q2314019; PIDN:AAD07934.1; PID:q2314022; TIGR:HB

A:Experimental source: strain 26695

R.Cover, T.L.; Tummuru, M.K.R.; Cao, P.; Thompson, S.A.; Blaser, M.J.

J. Biol. Chem. 269, 10566-10573, 1994

A:Title: Divergence of genetic sequences for the vacuolating cytotoxin among *Helicob*

A:Reference number: A53739; MUID:94193753; PMID:8144644

A:Accession: C53739

A:Molecule type: DNA

A:Residues: 278-279, 'I', 281-331, 'N', 333-391, 'V', 393-431, 'Q', 433-439, 'H', 441-465 <COV

A:Cross-references: GB:U05676; NID:q471727; PIDN:AAI17656.1; PID:q471728

A:Experimental source: strain ATCC 49503

C:Superfamily: cysteine-tRNA ligase

C:Keywords: aminocacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 83.3%; Score 30; DB 1; Length 465;  
Best Local Similarity 75.0%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSIERFARL 8  
|||||  
DB 159 SSIERFARL 166

RESULT 14

D71884

cysteine-tRNA ligase (EC 6.1.1.16) - *Helicobacter pylori* (strain J99)

C:Species: *Helicobacter pylori*

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 03-Jun-2002

C:Accession: D71884

R.Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.)

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric I

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: D71884

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-465 <ARN>

A:Cross-references: GB:AE001511; GB:AE001439; NID:q4155382; PIDN:AAD06399.1; PID:q41

A:Experimental source: strain J99

C:Genetics:

A:Gene: cyst

C:Superfamily: cysteine-tRNA ligase

C:Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 83.3%; Score 30; DB 2; Length 465;  
Best Local Similarity 62.5%; Pred. No. 32;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIERFARL 8  
|||||  
DB 159 SSIERFARL 166

RESULT 15

V0680H

glycoprotein B precursor - equine herpesvirus 4 (strain 1942)

C:Species: equine herpesvirus 4

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999

C:Accession: A31860

R.Rigbio, M.P.; Cullinane, A.A.; Onions, D.E.

J. Virol. 63, 1123-1133, 1989

A:Title: Identification and nucleotide sequence of the glycoprotein gb gene of equin

A;Reference number: A31880; MUID:89125704; PMID:2915378  
A;Accession: A31880  
A;Molecule type: DNA  
A;Residues: 1-919 <RIG>  
A;Cross-references: GB:M26171; NID:g341446; PIDN:AAA46106.1; PID:g514920  
C;Superfamily: herpesvirus glycoprotein B  
C;Keywords: glycoprotein; transmembrane protein  
F;1-28/Domain: signal sequence #status predicted <SIG>  
F;29-919/Product: glycoprotein B #status predicted <GPB>  
F;740-809/Domain: transmembrane #status predicted <TMN>  
F;106,216,321,364,438,456,493,499,666,688/Binding site: carbohydrate (asn) (covalent) #  
Query Match 83.3%; Score 30; DB 1; Length 919;  
Best Local Similarity 87.5%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SSIEFARL 8  
Db 512 SSIEFARL 519

Search completed: May 22, 2003, 12:10:11  
Job time : 16 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 22, 2003, 12:01:57 ; Search time 7.5 Seconds

(without alignments)  
44,241 Million cell updates/sec

Title: US-09-719-494-10

Perfect score: 36

Sequence: 1 SSIEFARL 8

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	903	1	VGLB_HSV1F
2	36	100.0	904	1	VGLB_HSV1F
3	36	100.0	904	1	VGLB_HSV1K
4	36	100.0	904	1	VGLB_HSV1P
5	36	100.0	904	1	VGLB_HSV23
6	36	100.0	904	1	VGLB_HSV2H
7	35	97.2	885	1	VGLB_HSV2S
8	35	97.2	917	1	VGLB_HSVB2
9	31	86.1	920	1	VGLB_HSVSM
10	30	83.3	344	1	ARAL_YEAST
11	30	83.3	435	1	ERPL_TETTH
12	30	83.3	445	1	ERPL_TETTH
13	30	83.3	445	1	ERPL_OXYTR
14	30	83.3	445	1	ERPL_STYLE
15	30	83.3	457	1	ERPL_STYMT
16	30	83.3	465	1	ERPL_GIALA
17	30	83.3	465	1	STC_HELPJ
18	30	83.3	499	1	STC_HELPJ
19	30	83.3	919	1	VGLB_HSVB4
20	30	83.3	979	1	VGLB_HSVEL
21	30	83.3	980	1	VGLB_HSVEL
22	30	83.3	980	1	VGLB_HSVB4
23	29	80.6	154	1	VGLB_HSVB4
24	29	80.6	288	1	REPA_ECOLI
25	29	80.6	317	1	KDR2_SALTI
26	29	80.6	868	1	VGLB_YZVD
27	29	80.6	953	1	LKA3_PASHA
28	29	80.6	955	1	LKA3_PASHA
29	29	80.6	1344	1	IF3A_MOUSE
30	29	80.6	1382	1	IF3A_MOUSE
31	28	77.8	250	1	PHON_SALTY
32	28	77.8	250	1	PHON_SALTY
33	28	77.8	256	1	PCNA_NPVAC

34	28	77.8	304	1	NOD1_RHIS3
35	28	77.8	306	1	NOD1_BRAVA
36	28	77.8	311	1	NOD1_RHILV
37	28	77.8	312	1	VPR_BPMU
38	28	77.8	340	1	NOD1_RHILV
39	28	77.8	355	1	NOD1_RHIME
40	28	77.8	358	1	RBB5_HUMAN
41	28	77.8	552	1	SYO_CLOPE
42	28	77.8	660	1	ALIA_STRPN
43	28	77.8	750	1	CBAB_BACTV
44	28	77.8	933	1	VGLB_HSV1
45	28	77.8	4466	1	DYHC_ANTCR

## ALIGNMENTS

RESULT 1	VGLB_HSV1F	STANDARD;	PRT;	903 AA.
ID	VGLB_HSV1F			
AC	P06436;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Glycoprotein B precursor.			
GN	GB OR UL27.			
OS	Herpes simplex virus (type 1 / strain F).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Alphaherpesvirinae; Simplexvirus.			
OX	NCBI_TaxID=10304;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85083254; PubMed=2981343;			
RA	Pellet P.E., Kousoulas K.G., Pereira L., Rolzman B.;			
RT	"Anatomy of the herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody-resistant mutants."			
RL	J. Virol. 53:243-253(1985).			
RN	[2]			
RP	SEQUENCE OF 1-176 FROM N.A.			
RX	MEDLINE=88306232; PubMed=2457278;			
RA	Hammer Schmidt W., Contraths F., Manhertz J., Buhk H.-J., Paul G., Ludwig H.;			
RT	"Common epitopes of glycoprotein B map within the major DNA-binding proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1)."			
RL	Virolgy 165:406-418(1988).			
CC	-1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.			
CC	-1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH, GB, GC, GD, GI, AND GE.			
CC	-1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE REQUIRED FOR VIRAL GROWTH.			
CC	-1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).			
CC	EMBL: M14164; AAA45776.1; -			
CC	EMBL: M21633; AAA45788.1; -			
DR	PIR: A03750; VGBB1.			
DR	InterPro: IPR000234; Glycoprot_B.			
DR	Pfam: PF00606; Glycoprotein_B.1.			
DR	ProDom: PD000693; Glycoprot_B.1.			
FT	Glycoprotein; Transmembrane; Signal.			
FT	SIGNAL.			
FT	CHAIN			
FT	DOMAIN			
FT	TRANSMEM			
FT	730			
FT	745			
FT	POTENTIAL.			
FT	GLYCOPROTEIN B.			
FT	EXTRACELLULAR (POTENTIAL).			

```

FT TRANSMEM 751 770 POTENTIAL.
FT TRANSEM 774 794 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 795 903
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 673 673 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100104 MM; 73BDCN7813DB35E8 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 1; Length 903;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8
DB 498 SSIFFARL 505

RESULT 2
VGLB_HSV11 STANDARD; PRT; 904 AA.
ID VGLB_HSV11
AC P10211;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein B precursor.
GN GB OR UL27.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Simplexvirus.
NCBI_TaxID=10299;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=88274327; PubMed=2839594;
RX McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McAb D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).

[2]
RN -1 SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC -1 MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC GB, GC, GG, GD, GI, AND GE.
CC -1 MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC REQUIRED FOR VIRAL GROWTH.
CC -1 SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10879; BAA01673.1; -
DR EMBL: X14112; CAA32320.1; -
DR PIR: I30084; VGBEW7.
DR InterPro: IPR000234; Glycoprot_B.
DR Pfam: PF00606; Glycoprotein_B.1.
DR ProDom: PD000693; Glycoprot_B.1.
DR Glycoprotein; Transmembrane; Signal.
KW SIGNAL
FT CHAIN 1 30
FT DOMAIN 31 904 GLYCOPROTEIN B.
FT TRANSEM 31 730 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 731 746 POTENTIAL.
FT TRANSEM 752 771 POTENTIAL.
FT TRANSEM 775 795 POTENTIAL.
FT DOMAIN 796 904 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 904 AA; 100292 MM; 2C14E8B1284C1E3A CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 1; Length 904;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8
DB 499 SSIFFARL 506

RESULT 3
VGLB_HSV1K STANDARD; PRT; 904 AA.
ID VGLB_HSV1K
AC P06437;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycoprotein B precursor.
GN GB OR UL27.
OS Herpes simplex virus (type 1 / strain KOS).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Simplexvirus.
NCBI_TaxID=10306;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=84174058; PubMed=6324454;
RX Bzik D.J., Fox B.A., Deluca N.A., Person S.;
RA "Nucleotide sequence specifying the glycoprotein gene, GB, of herpes
RT simplex virus type 1."
RL Virology 133:301-314(1984).

[2]
RN REVISIONS.
RP MEDLINE=87071654; PubMed=3024391;
RX Bzik D.J., Debroy C., Fox B.A., Pederson N.E., Person S.;
RA "The nucleotide sequence of the gb glycoprotein gene of HSV-2 and
RT comparison with the corresponding gene of HSV-1."
RL Virology 155:322-333(1986).

[3]
RN REVISED.
RP Pederson N.E.;
RA submitted (Apr-1987) to the EMBL/GenBank/DBJ databases.
RL -1 SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.
CC -1 MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,
CC GB, GC, GG, GD, GI, AND GE.
CC -1 MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE
CC REQUIRED FOR VIRAL GROWTH.
CC -1 SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01760; AAA45774.1; -
DR PIR: A03751; VGBEKL.
DR InterPro: IPR000234; Glycoprot_B.
DR Pfam: PF00606; Glycoprotein_B.1.
DR ProDom: PD000693; Glycoprot_B.1.
DR Glycoprotein; Transmembrane; Signal.
KW SIGNAL
FT CHAIN 1 30
FT DOMAIN 31 904 GLYCOPROTEIN B.
FT TRANSEM 31 730 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 731 746 POTENTIAL.
FT TRANSEM 752 771 POTENTIAL.
FT TRANSEM 775 795 POTENTIAL.
FT DOMAIN 796 904 CYTOPLASMIC (POTENTIAL).

```

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8  
 |||||

DB 499 SSIFFARL 506

RESULT 4  
 VGLB\_HSV1P STANDARD; PRT; 904 AA.

AC P08655;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB OR UL27.  
 OS Herpes simplex virus (type 1 / strain Patton).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=103108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87112925; PubMed=3027364;  
 RA Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,  
 RA Burke R.L.;  
 RT "Structure and expression of the herpes simplex virus type 2  
 glycoprotein gb gene.";  
 RT J. Virol. 61:326-335(1987).  
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1: GH,  
 CC GB, GC, GG, GD, GI, AND GE.  
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: K03541; AAA45778.1; -  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR ProDom: PD000693; Glycoprot\_B; 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 904 GLYCOPROTEIN B.  
 FT DOMAIN 31 730 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 731 746 POTENTIAL.  
 FT TRANSMEM 752 772 POTENTIAL.  
 FT TRANSMEM 775 795 POTENTIAL.  
 FT DOMAIN 796 904 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 904 AA; 100115 MW; 7825E1DC830A626F CRC64;

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8  
 |||||

DB 499 SSIFFARL 506

RESULT 5  
 VGLB\_HSV23 STANDARD; PRT; 904 AA.

AC P06763;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB OR UL27 OR GB2.  
 OS Herpes simplex virus (type 2 / strain 333).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87112925; PubMed=3027364;  
 RA Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,  
 RA Burke R.L.;  
 RT "Structure and expression of the herpes simplex virus type 2  
 glycoprotein gb gene.";  
 RT J. Virol. 61:326-335(1987).  
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND  
 CC 2: GH, GB, GC, GG, GD, GI, AND GE.  
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M1518; AAA45837.1; -  
 DR PIR: A26790; VGBB82.  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR ProDom: PD000693; Glycoprot\_B; 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 904 GLYCOPROTEIN B.  
 FT DOMAIN 23 727 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 728 743 POTENTIAL.  
 FT TRANSMEM 749 768 POTENTIAL.  
 FT TRANSMEM 772 792 POTENTIAL.  
 FT DOMAIN 793 904 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 904 AA; 100186 MW; A8B36F74FDBC8539 CRC64;

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIFFARL 8  
 |||||

DB 496 SSIFFARL 503



## RESULT 6

VGIB\_HSV2H STANDARD; PRT; 904 AA.  
 AC P08666; P89450;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB OR UL27.  
 OS Herpes simplex virus (type 2 / strain HGS2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 NCBI\_TaxID=10315;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87071654; PubMed=3024391;  
 RA Bazik D.J., DeBrooy C., Fox B.A., Pederson N.E., Person S.;  
 RT "The nucleotide sequence of the gb glycoprotein gene of HSV-2 and  
 comparison with the corresponding gene of HSV-1.";  
 RN Virology 135:322-333(1986).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Dolan A.;  
 RT Submitted (FEB-1997) to the EMBL/Genbank/DDJ databases.  
 CC -1 SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -1 MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND  
 CC 2: GH, GB, GC, GG, GD, GI, AND GE.  
 CC -1 MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -1 SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M14923; AA66440.1; -  
 DR EMBL: 286099; CAB06752.1; -  
 DR PIR: A25611; VGBEK2.  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR ProDom: PD000693; Glycoprot\_B; 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 904  
 FT DOMAIN 23 727  
 FT TRANSMEM 728 743  
 FT TRANSMEM 749 768  
 FT TRANSMEM 772 792  
 FT DOMAIN 793 904  
 FT CARBOHYD 82 82  
 FT CARBOHYD 136 136  
 FT CARBOHYD 393 393  
 FT CARBOHYD 425 425  
 FT CARBOHYD 466 466  
 FT CARBOHYD 671 671  
 FT CONFLICT 92 92  
 FT CONFLICT 198 198  
 FT CONFLICT 308 308  
 FT CONFLICT 438 438  
 FT CONFLICT 568 568  
 FT CONFLICT 619 620  
 FT CONFLICT 636 636  
 SQ SEQUENCE 904 AA; 100217 MW; AB050A3AFB4F1066 CRC64;

Query Match 100.0%; Score 36; DB 1; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSTEFAFL 8  
 DB 496 SSTEFAFL 503

## RESULT 7

VGIB\_HSV2S STANDARD; PRT; 885 AA.  
 AC P24994;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB OR UL27.  
 OS Herpes simplex virus (type 2 / strain SA8) (Simian agent 8).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 NCBI\_TaxID=10316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91374035; PubMed=1895066;  
 RC STRAIN-B264;  
 RA Borchers K., Weigelt W., Buhk H.-J., Ludvig H., Mankertz J.;  
 RT "Conserved domains of glycoprotein B (gb) of the monkey virus, simian  
 agent 8, identified by comparison with herpesvirus gbs.";  
 RN J. Gen. Virol. 72:2299-2304(1991).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Dolan A.;  
 RT Submitted (FEB-1997) to the EMBL/Genbank/DDJ databases.  
 CC -1 SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -1 MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND  
 CC 2: GH, GB, GC, GG, GD, GI, AND GE.  
 CC -1 MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -1 SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X56935; CAA40256.1; -  
 DR PIR: J01332; VGBESA.  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR ProDom: PD000693; Glycoprot\_B; 1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 34  
 FT CHAIN 35 885  
 FT DOMAIN 35 715  
 FT TRANSMEM 716 731  
 FT TRANSMEM 737 756  
 FT TRANSMEM 760 780  
 FT DOMAIN 781 885  
 FT CARBOHYD 68 68  
 FT CARBOHYD 122 122  
 FT CARBOHYD 379 379  
 FT CARBOHYD 411 411  
 FT CARBOHYD 659 659  
 SQ SEQUENCE 885 AA; 97811 MW; 39E495B329AB94E4 CRC64;

Query Match 97.2%; Score 35; DB 1; Length 885;  
 Best Local Similarity 87.5%; Pred. No. 1.7;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSTEFAFL 8  
 DB 484 SSTEFAFL 491

## RESULT 8

VGIB\_HSVB2

ID VGLB\_HSVB2 STANDARD; PRT; 917 AA.  
 AC P12641;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B-1 precursor.  
 OS Bovine herpesvirus type 2 (strain BMV) (Bovine gammaherpesvirus 2).  
 OC Alphaherpesvirinae; Simplexvirus.  
 NCBI\_TaxID:10296;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:88306231; PubMed:2841793;  
 RA Hammerschmidt W., Contraths F., Mankertz J., Pauli G., Ludwig H.,  
 RA Buhk H.-J.;  
 RT "Conservation of a gene cluster including glycoprotein B in bovine  
 RT herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";  
 RL Virology 165:388-405(1988).  
 RN [2]  
 RP SEQUENCE OF 1-200 FROM N.A.  
 RX MEDLINE:88306232; PubMed:2457278;  
 RA Hammerschmidt W., Contraths F., Mankertz J., Buhk H.-J., Pauli G.,  
 RA Ludwig H.;  
 RT "Common epitopes of glycoprotein B map within the major DNA-binding  
 RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex  
 RT virus type 1 (HSV-1).";  
 RL Virology 165:406-418(1988).  
 CC -1- FUNCTION: GB1 IS A 130 kDa GLYCOPROTEIN WHICH IS NECESSARY FOR THE  
 CC PENETRATION OF THE VIRUS INTO THE HOST CELL AND THE INDUCTION OF A  
 CC SYNCYTIAL PHENOTYPE.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M21628; AAA46053.1; -;  
 DR EMBL: M21632; AAA46052.1; -;  
 DR PIR: C29242; VGBEBH.  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B.1.  
 DR Prodom: PD000693; Glycoprot\_B.1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 917 GLYCOPROTEIN B-1.  
 FT TRANSMEM 578 594 POTENTIAL.  
 FT TRANSMEM 770 786 POTENTIAL.  
 FT TRANSMEM 795 811 POTENTIAL.  
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 505 505 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 692 692 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 917 917 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 917 AA; 101882 MW; 1B96CBF50DB4D3FC CRC64;

Query Match 97.2%; Score 35; DB 1; Length 917;  
 Best Local Similarity 87.5%; Pred. No. 1.8;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIEFARL 8  
 Db 517 SVEFARL 524

RESULT 9

VGLB\_HSVSM STANDARD; PRT; 920 AA.  
 ID VGLB\_HSVSM  
 AC 004464;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein B precursor.  
 OS Herpesvirus salm1r1 (type 1 / strain MV-5-4-PSI) (Marmoset  
 OS herpesvirus).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae.  
 NCBI\_TaxID:10353;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:9328440; PubMed:8385913;  
 RA Eberle R., Black D.;  
 RT "Sequence analysis of herpes simplex virus gb gene homologs of two  
 RT platyrrhine monkey alpha-herpesviruses.";  
 RL Arch. Virol. 129:167-182(1993).  
 CC -1- SUBUNIT: DIMER, PROBABLY LINKED BY DISULFIDE BONDS.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- MISCELLANEOUS: GB IS THE ONLY GLYCOPROTEIN THAT IS KNOWN TO BE  
 CC REQUIRED FOR VIRAL GROWTH.  
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN B FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M95786; AAA43841.1; -;  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B.1.  
 DR Prodom: PD000693; Glycoprot\_B.1.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 920 POTENTIAL.  
 FT DOMAIN 29 739 GLYCOPROTEIN B.  
 FT TRANSMEM 740 755 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 761 781 POTENTIAL.  
 FT TRANSMEM 784 804 POTENTIAL.  
 FT DOMAIN 805 920 POTENTIAL.  
 FT CARBOHYD 98 98 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 733 733 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 920 AA; 101661 MW; 62C4D0315BDC2DB7 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 920;  
 Best Local Similarity 85.7%; Pred. No. 17;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SIEFARL 8  
 Db 509 SVEFARL 515

RESULT 10  
 ARA1\_YEAST STANDARD; PRT; 344 AA.  
 ID ARA1\_YEAST  
 AC P38115;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE D-arabinose dehydrogenase [NAD(P)+] heavy chain (EC 1.1.1.117).  
 GN ARA1 OR YBR149W OR YBR1177.

RESULT 10

05 Saccharomyces cerevisiae (Baker's yeast).  
 06 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 07 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 08 NCBI\_TaxID=4932;  
 09 (1)  
 10 SEQUENCE FROM N.A.  
 11 STRAIN-S288C;  
 12 Eutlian K.-D., Koetter P., Rose M., Becker J., Grey M., Li Z.,  
 13 Niegemann E., Schenk-Greeninger R., Servos J., Wehner E.,  
 14 Wolter R., Brendel M., Bauer J., Braun H., Dorn K., Duesterhus S.,  
 15 Gruenbein R., Hedges D., Klesau P., Korol S., Krems B., Proft M.,  
 16 Stegers K., Baur A., Boles E., Miosga T.,  
 17 Schaff-Gerstenschlaeger I., Zimmermann F.K.;  
 18 Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.  
 19 (2)  
 20 SEQUENCE FROM N.A.  
 21 MEDLINE-91171289; PubMed-2005616;  
 22 Martinez-Soriano J.P., Wong W.M., van Ryk D.I., Nazari R.N.;  
 23 "A widely distributed 'CAT' family of repetitive DNA sequences,"  
 24 J. Mol. Biol. 217:629-635(1991).  
 25 (3)  
 26 CHARACTERIZATION, AND SEQUENCE OF 7-20.  
 27 MEDLINE-9117109; PubMed-9920381;  
 28 Kim S.T., Huh W.K., Lee B.H., Kang S.O.;  
 29 "D-arabinose dehydrogenase and its gene from Saccharomyces  
 30 cerevisiae,"  
 31 Biochim. Biophys. Acta 1429:29-39(1998).  
 32 -1- FUNCTION: CATALYZES THE OXIDATION OF D-ARABINOSE, L-XYLOSE, L-  
 33 FUCCOSE AND L-GALACTOSE IN THE PRESENCE OF NADP+.  
 34 -1- CATALYTIC ACTIVITY: D-arabinose + NAD(P)(+) = D-arabinono-1,4-  
 35 lactone + NAD(P)H.  
 36 -1- SUBUNIT: HETERODIMER OF A HEAVY CHAIN AND A LIGHT CHAIN.  
 37 -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 38 -1- MISCELLANEOUS: EXHIBITS MAXIMUM ACTIVITY AT PH 10.0 AND AROUND 30  
 39 DEGREES CELSIUS.  
 40 -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE FAMILY.  
 41 -----  
 42 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 43 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 44 the European Bioinformatics Institute. There are no restrictions on its  
 45 use by non-profit institutions as long as its content is in no way  
 46 modified and this statement is not removed. Usage by and for commercial  
 47 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 48 or send an email to license@sib-sib.ch).  
 49 -----  
 50 CC EMBL: Z36018; CAAB5107.1; -  
 51 DR EMBL: M95580; AAA35037.1; -  
 52 DR PIR: S46020; S46020.  
 53 DR HSP: P06632; LHM6.  
 54 DR SGD: S0000353; AAP1.  
 55 DR InterPro: IPR001395; Aldo/ket\_red.  
 56 DR Pfam: PF00248; Aldo\_ket\_red.1.  
 57 DR PRINTS: PR00069; ALDKETREDTASE.  
 58 DR ProDom: PD000288; Aldo/ket\_red.1.  
 59 DR PROSITE: PS00798; ALDOKETO\_REDUCTASE\_1; 1.  
 60 DR PROSITE: PS00062; ALDOKETO\_REDUCTASE\_2; 1.  
 61 DR PROSITE: PS00063; ALDOKETO\_REDUCTASE\_3; 1.  
 62 KM Oxidoreductase; NAD.  
 63 ACT\_SITE 71  
 64 FT CONFLICT 56 62 AAIRAGY -> LOSKLDN (IN REF. 2).  
 65 FT CONFLICT 71 73 YET -> SR (IN REF. 2).  
 66 SQ SEQUENCE 344 AA; 38883 MW; F9B6D9333B18FEBC CRC64;

Query Match 83.3%; Score 30; DB 1; Length 344;  
 Best Local Similarity 87.5%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSTEFAFL 8  
 DB 294 SSTEFAFL 301

RESULT 11

ERF1\_TETH STANDARD; PRT; 435 AA.  
 ID ERF1\_TETH  
 AC 09U805;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Eukaryotic peptide chain release factor subunit 1 (ERF1) (Eukaryotic  
 release factor 1).  
 GN ERF1.  
 OS Tetrahymena thermophila.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 CC Tetrahymenina; Tetrahymena.  
 OX NCBI\_TaxID=5911;  
 (1)  
 10 SEQUENCE FROM N.A.  
 11 MEDLINE-99402978; PubMed-10471834;  
 12 Karamychev A.L., Ito K., Nakamura Y.;  
 13 "Cloning, purification and complex formation with yeast erf3,"  
 14 FEBS Lett. 457:483-488(1999).  
 15 (2)  
 16 SEQUENCE FROM N.A.  
 17 MEDLINE-2113188; PubMed-11160924;  
 18 Inagaki Y., Doi Little W.F.;  
 19 "Class I release factors in ciliates variant genetic codes,"  
 20 Nucleic Acids Res. 29:921-927(2001).  
 21 -1- FUNCTION: Directs the termination of nascent peptide synthesis  
 22 (translocation) in response to the termination codon UGA.  
 23 In T. thermophila UAA and UAG codes for glutamine.  
 24 -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.  
 25 -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 26 -1- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.  
 27 -----  
 28 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 29 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 30 the European Bioinformatics Institute. There are no restrictions on its  
 31 use by non-profit institutions as long as its content is in no way  
 32 modified and this statement is not removed. Usage by and for commercial  
 33 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 34 or send an email to license@sib-sib.ch).  
 35 -----  
 36 CC EMBL: AB026195; BAA85336.1; -  
 37 DR EMBL: AF298833; AAK07831.1; -  
 38 DR HSP: P46055; IDT9.  
 39 DR InterPro: IPR004403; ERF1.  
 40 DR InterPro: IPR005140; erf1\_1.  
 41 DR InterPro: IPR005141; erf1\_2.  
 42 DR InterPro: IPR005142; erf1\_3.  
 43 DR Pfam: PF03463; erf1\_1; 1.  
 44 DR Pfam: PF03464; erf1\_2; 1.  
 45 DR Pfam: PF03465; erf1\_3; 1.  
 46 DR TIGRFAMs: TIGR00108; erf; 1.  
 47 KM Protein biosynthesis.  
 48 SQ SEQUENCE 435 AA; 49558 MW; 89013ED4C8612646 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 435;  
 Best Local Similarity 75.0%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SSTEFAFL 8  
 DB 185 SSTEFAFL 192

RESULT 12

ERF1\_OYTR STANDARD; PRT; 445 AA.

AC 09BKA3; 09BKA2; (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Eukaryotic peptide chain release factor subunit 1 (ERF1) (Eukaryotic  
 release factor 1).

```

GN ERF1.
OS Oxytricha trifallax.
OC Eukaryota, Alveolata, Ciliophora, Spirotrichea, Stichotrichia;
OC Stichotrichidae; Oxytrichidae; Oxytricha.
RN NCBI_TaxID=5946;
RP SEQUENCE FROM N.A.
RA MEDLINE=2113168; PubMed=11160924;
RA Nagaki Y., Doellittle W.F.;
RT Class I release factors in ciliates variant genetic codes.";
RL Nucleic Acids Res. 29:921-927(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21154807; PubMed=11231122;
RX Lozupone C.A., Knight R.D., Landweber L.F.;
RT "The molecular basis of nuclear genetic code change in ciliates.";
RL Curr. Biol. 11:65-74(2001).
CC -1- FUNCTION: Directs the termination of nascent peptide synthesis
CC (translation) in response to the termination codon UGA.
CC In O. trifallax UGA and UAG codes for glutamine.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF298830; AAK07828.1;
DR EMBL: AF317832; AAK12090.1;
DR HSSP: P46055; IPT9.
DR InterPro: IPR004403; eRF1.
DR InterPro: IPR005140; eRF1_1.
DR InterPro: IPR005141; eRF1_2.
DR InterPro: IPR005142; eRF1_3.
DR Pfam: PF03463; eRF1_1; 1.
DR Pfam: PF03464; eRF1_2; 1.
DR Pfam: PF03465; eRF1_3; 1.
DR TIGRFAMs: TIGR00108; eRF; 1.
DR TIGRFAMs: TIGR00108; eRF; 1.
KW Protein biosynthesis.
SQ SEQUENCE 445 AA; 49716 MW; 7D637F4C2572A09F CRC64;

Query Match 83.3%; Score 30; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSIERFARL 8
DB 191 SSVFARL 198

RESULT 13
ERF1_STYLE STANDARD; PRT; 445 AA.
AC Q9BMMO;
ID ERF1_STYLE
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Eukaryotic peptide chain release factor subunit 1 (erf1) (Eukaryotic
DE release factor 1).
GN ERF1.
OS Stylonychia lemnae.
OC Eukaryota, Alveolata, Ciliophora, Spirotrichea, Stichotrichia;
OC Stichotrichidae; Oxytrichidae; Stylonychia.
OX NCBI_TaxID=5949;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21154807; PubMed=11231122;
RX Lozupone C.A., Knight R.D., Landweber L.F.;

```

```

RT "The molecular basis of nuclear genetic code change in ciliates.";
RL Curr. Biol. 11:65-74(2001).
CC -1- FUNCTION: Directs the termination of nascent peptide synthesis
CC (translation) in response to the termination codon UGA.
CC In Stylonychia UGA and UAG codes for glutamine.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF317834; AAK12092.1;
DR HSSP: P46055; IPT9.
DR InterPro: IPR004403; eRF1.
DR InterPro: IPR005140; eRF1_1.
DR InterPro: IPR005141; eRF1_2.
DR InterPro: IPR005142; eRF1_3.
DR Pfam: PF03463; eRF1_1; 1.
DR Pfam: PF03464; eRF1_2; 1.
DR Pfam: PF03465; eRF1_3; 1.
DR TIGRFAMs: TIGR00108; eRF; 1.
KW Protein biosynthesis.
SQ SEQUENCE 445 AA; 49558 MW; CCE0D191E15D74BC CRC64;

Query Match 83.3%; Score 30; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSIERFARL 8
DB 191 SSVFARL 198

RESULT 14
ERF1_STYLE STANDARD; PRT; 445 AA.
AC Q9BMMO;
ID ERF1_STYLE
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Eukaryotic peptide chain release factor subunit 1 (erf1) (Eukaryotic
DE release factor 1).
GN ERF1.
OS Stylonychia mytilus.
OC Eukaryota, Alveolata, Ciliophora, Spirotrichea, Stichotrichia;
OC Stichotrichidae; Oxytrichidae; Stylonychia.
OX NCBI_TaxID=5950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21154807; PubMed=11231122;
RX Lozupone C.A., Knight R.D., Landweber L.F.;
RT "The molecular basis of nuclear genetic code change in ciliates.";
RL Curr. Biol. 11:65-74(2001).
CC -1- FUNCTION: Directs the termination of nascent peptide synthesis
CC (translation) in response to the termination codon UGA.
CC In Stylonychia UGA and UAG codes for glutamine.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC -----
DR EMBL: AF117833; AAK12091.1; -
DR HSSP: P46055; 1DT9;
DR InterPro: IPR004403; ERF1_1
DR InterPro: IPR005140; ERF1_1
DR InterPro: IPR005141; ERF1_2
DR InterPro: IPR005142; ERF1_3
DR Pfam: PF03463; ERF1_1; 1
DR Pfam: PF03464; ERF1_2; 1
DR Pfam: PF03465; ERF1_3; 1
DR TIGRPFAM: TIGR00108; ERF; 1
DR Protein biosynthesis.
SQ SEQUENCE 445 AA; 49670 MW; 0F362FEA1BD50B8 CRC64;

```

```

Query Match      83.3%; Score 30; DB 1; Length 445;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 SSIEFARL 8
   ||: ||||
DB 191 SSVRFARL 198

```

```

RESULT 15
ERF1_GIALA STANDARD; PRT; 457 AA.
AC Q9NCP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Eukaryotic peptide chain release factor subunit 1 (ERF1) (Eukaryotic
DE release factor 1).
GN ERF1.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20293306; PubMed=10833194;
RA Inagaki Y., Doolittle W.F.;
RT Evolution of the eukaryotic translation termination system: origins
RT of release factors."
RL Mol. Biol. Evol. 17:882-889(2000).
CC -1- FUNCTION: Directs the termination of nascent peptide synthesis
CC (translation) in response to the termination codons UAA, UAG and
CC UGA (By similarity).
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, ONE OF WHICH BINDS GTP.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC RELEASE FACTOR 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: AF198107; AAF74402.1; -
DR HSSP: P46055; 1DT9;
DR InterPro: IPR004403; ERF1_1
DR InterPro: IPR005140; ERF1_1
DR InterPro: IPR005141; ERF1_2
DR InterPro: IPR005142; ERF1_3
DR Pfam: PF03463; ERF1_1; 1
DR Pfam: PF03464; ERF1_2; 1
DR Pfam: PF03465; ERF1_3; 1
DR TIGRPFAM: TIGR00108; ERF; 1
DR Protein biosynthesis.
KW SEQUENCE 457 AA; 51029 MW; EE185FFFD1F0C943 CRC64;

```

```

Query Match      83.3%; Score 30; DB 1; Length 457;
Best Local Similarity 75.0%; Pred. No. 14;

```

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 SSIEFARL 8
   ||: ||||
DB 191 SSVRFARL 198

```

Search completed: May 22, 2003, 12:08:39  
Job time : 8.5 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:04:33 ; Search time 25 Seconds  
(without alignments)  
65.935 Million cell updates/sec

Title: US-09-719-494-10  
Perfect score: 36  
Sequence: 1 SSIEFARL 8

Scoring table: BLOSUM62  
Gapop: 10.0 , Gapext: 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	583	12 068981	068981 human herpe
2	36	100.0	887	12 09YW93	09YW93 macropodid
3	36	100.0	893	12 065540	065540 baboon herp
4	36	100.0	894	12 09YW92	09YW92 macropodid
5	36	100.0	901	12 069464	069464 herpes simp
6	36	100.0	901	12 069465	069465 herpes simp
7	36	100.0	902	12 069095	069095 herpes simp
8	36	100.0	903	12 069076	069076 human herpe
9	36	100.0	904	12 069526	069526 human herpe
10	36	100.0	904	12 09Q1M8	09Q1M8 human herpe
11	36	100.0	904	12 091WU4	091WU4 human herpe
12	36	100.0	904	12 09DXE4	09DXE4 herpes simp
13	36	100.0	904	12 037453	037453 herpes simp
14	36	100.0	904	12 091C63	091C63 herpes simp
15	36	100.0	904	12 089920	089920 herpes simp
16	35	97.2	885	12 069387	069387 cercopithec

17	35	97.2	891	12 066018	066018 simian herp
18	35	97.2	891	12 09J052	09J052 simian herp
19	35	97.2	908	12 065538	065538 baboon herp
20	31	86.1	261	17 08RTJ6	08RTJ6 methanosarc
21	31	86.1	429	5 08W073	08W073 trichomonas
22	30	83.3	83	16 09X8K3	09X8K3 streptomyce
23	30	83.3	153	16 09PJ01	09PJ01 campylobact
24	30	83.3	350	5 019582	019582 caenorhabdi
25	30	83.3	386	5 08SRZ6	08SRZ6 encephalito
26	30	83.3	389	17 08RTQ40	08RTQ40 methanosarc
27	30	83.3	406	2 09JCM9	09JCM9 helicobacte
28	30	83.3	437	5 0965E7	0965E7 paramecium
29	30	83.3	437	5 0965E6	0965E6 paramecium
30	30	83.3	502	5 08TR13	08TR13 dictyosteli
31	30	83.3	586	3 09HG02	09HG02 emericella
32	30	83.3	950	12 066678	066678 equine herp
33	30	83.3	975	12 039275	039275 equine herp
34	30	83.3	980	12 066682	066682 equine herp
35	30	83.3	1315	3 09GRX8	09GRX8 schizosacch
36	29	80.6	208	2 051865	051865 pasteurella
37	29	80.6	233	2 051865	051865 pasteurella
38	29	80.6	263	2 09K1W5	09K1W5 thioabaciliu
39	29	80.6	288	2 052345	052345 unclassified
40	29	80.6	318	16 P93279	P93279 mycobacteri
41	29	80.6	321	5 09XWD4	09XWD4 caenorhabdi
42	29	80.6	415	2 024782	024782 pseudomonas
43	29	80.6	452	16 08UJH1	08UJH1 agrobacteri
44	29	80.6	478	2 054231	054231 streptococc
45	29	80.6	532	16 09K6D7	09K6D7 bacillus ha

## ALIGNMENTS

## RESULT 1

ID 068981

PRELIMINARY; PRT; 583 AA.

AC 068981;

01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HSV1 (Mutant strain tsb5), glycoprotein B (gb) gene (Fragment).

human herpesvirus 1.

OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Alphaherpesvirinae; Simplexvirus.

OC NCBI\_TaxID=10298;

[1]

RP SEQUENCE FROM N.A.

MEDLINE=84301857; PubMed=6089415;

RA Bzik D.J., Fox B.A., Deluca N.A., Person S.;

\*Nucleotide sequence of a region of the herpes simplex virus type 1 gb

RT glycoprotein gene: Mutations affecting rate of virus entry and cell

fusion.

RL Virology 137:185-190(1984).

EMBL; K02720; AAA5777.1; -.

DR InterPro: IPR000234; Glycoprot\_B.

Pfam: PF00606; Glycoprotein\_B; 1.

DR Pfam: PD000693; Glycoprot\_B; 1.

NON\_TER

FT SEQUENCE

583 AA; 64914 MW; 9F4997AFCC96E457 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 583;  
Best local Similarity 100.0%; Pred. No. 4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSIEFARL 8

|||||||

Db 178 SSIEFARL 185

RESULT 2

Q9YW93

PRELIMINARY;

PRT;

887 AA.

```

AC Q9Y93;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Glycoprotein B.
GN UL27.
OS Macropodid herpesvirus type 1 (parma wallaby herpesvirus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
OX NCBI_TaxID=83441;
RN [1]
RP MEDLINE=99171712; PubMed=10073704;
RX Mahony T.J., Smith G.A., Thomson D.M.;
RT "Macropodid herpesviruses 1 and 2 occupy unexpected molecular
RT phylogenetic positions within the Alphaherpesvirinae.";
RL J. Gen. Virol. 80:433-436(1999).
DR EMBL; AF061754; AAD11960.1; -
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B.1.
DR Prodom; PD000693; Glycoprot_B.1.
SQ SEQUENCE 887 AA; 99739 MW; 0BE4FB4E098F128 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 887;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIEFARL 8
DB 482 SIEFARL 489

RESULT 3
ID 065540 PRELIMINARY; PRT; 893 AA.
AC Q65540;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GB glycoprotein.
GN UL27.
OS Baboon herpesvirus 2.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=36347;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X313;
RX MEDLINE=95251550; PubMed=7733825;
RT Eberle R., Black D.H., Lipper S., Hilliard J.K.;
RT "Herpesvirus papio 2, an SNA-like alpha-herpesvirus of baboons.";
RL Arch. Virol. 140:528-545(1995).
DR EMBL; U14663; AAB85650.1; -
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B.1.
DR Prodom; PD000693; Glycoprot_B.1.
SQ SEQUENCE 893 AA; 98520 MW; 06D9E973EE1712A1 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 893;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIEFARL 8
DB 487 SIEFARL 494

RESULT 4
OY92 PRELIMINARY; PRT; 894 AA.
AC Q9Y92;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Glycoprotein B.
GN UL27.
OS Macropodid herpesvirus type 2 (dorcopsis wallaby herpesvirus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
OX NCBI_TaxID=83440;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99171712; PubMed=10073704;
RX Mahony T.J., Smith G.A., Thomson D.M.;
RT "Macropodid herpesviruses 1 and 2 occupy unexpected molecular
RT phylogenetic positions within the Alphaherpesvirinae.";
RL J. Gen. Virol. 80:433-436(1999).
DR EMBL; AF061755; AAD11961.1; -
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B.1.
DR Prodom; PD000693; Glycoprot_B.1.
SQ SEQUENCE 894 AA; 100415 MW; 806CD45B72A8CB CRC64;

Query Match 100.0%; Score 36; DB 12; Length 894;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIEFARL 8
DB 485 SIEFARL 492

RESULT 5
ID 069464 PRELIMINARY; PRT; 901 AA.
AC Q69464;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Glycoprotein B precursor.
GN UL27.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CA4B;
RA Terhune S.S., Spear P.G.;
RT "Variability of herpes simplex virus glycoproteins in clinical
RT specimens.";
RT Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 80-901 FROM N.A.
RC STRAIN=CA4B;
RX MEDLINE=87112925; PubMed=3027364;
RX Stuve L.L., Brown-Shimer S., Pachl C., Najarian R., Dina D.,
RX Burke R.L.;
RT "Structure and expression of the herpes simplex virus type 2
RT glycoprotein gB gene.";
RL J. Virol. 61:326-335(1987).
DR EMBL; U12172; AAB60545.1; -
DR InterPro; IPR000234; Glycoprot_B.
DR Pfam; PF00606; Glycoprotein_B.1.
DR Prodom; PD000693; Glycoprot_B.1.
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT VARIANT 46 46 Q -> R (IN REF. 2).
FT VARIANT 57 57 R -> K (IN REF. 2).
FT VARIANT 327 327 Q -> R (IN REF. 2).
SQ SEQUENCE 901 AA; 99919 MW; C1AEBB4E000885E CRC64;

Query Match 100.0%; Score 36; DB 12; Length 901;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 SSIEFARL 8  
 DB 493 SSIEFARL 500

RESULT 6  
 ID 069465 PRELIMINARY; PRT; 901 AA.

AC 069465;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN UL27.  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10310;

RN (1)  
 RP SEQUENCE FROM N.A.

RA Terhune S.S., Spear P.G.;  
 RT "Variability of herpes simplex virus glycoproteins in clinical specimens."  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE OF 80-901 FROM N.A.

RC STRAIN-MMA;  
 RX MEDLINE=87112925; PubMed=3027364;

RA Stuve L.L., Brown-Shimer S., Paschl C., Najarian R., Dina D.,  
 Burke R.L.;

RT "Structure and expression of the herpes simplex virus type 2  
 glycoprotein gB gene."

RL J. Virol. 61:326-335(1987).  
 DR EMBL: 012174; AAB60547.1; -.

DR InterPro: IPR000234; Glycoprot.B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.

DR Prodom: PD000693; Glycoprot\_B; 1.  
 KW Signal.

FT SIGNAL 1 22 POTENTIAL.  
 FT VARIANT 57 57 R -> K (IN REF. 2).  
 FT VARIANT 68 68 E -> K (IN REF. 2).  
 FT VARIANT 327 327 E -> R (IN REF. 2).  
 FT VARIANT 393 393 E -> Q (IN REF. 2).

SO SEQUENCE 901 AA; 99949 MW; 6786F8DC747A4DF8 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 901;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 493 SSIEFARL 500

RESULT 7  
 ID 069095 PRELIMINARY; PRT; 902 AA.

AC 069095;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN GB2.  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10310;

RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN-333;  
 RX MEDLINE=88079667; PubMed=2446730;

RA Zwaagstra J.C., Leung W.C.;

RT "The nucleotide sequence of herpes simplex virus type 2 (333)  
 glycoprotein gB2 and analysis of predicted antigenic sites."  
 RL Can. J. Microbiol. 33:879-887(1987).  
 DR EMBL: M24771; AAB60540.1; -.

DR InterPro: IPR000234; Glycoprot.B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR Prodom: PD000693; Glycoprot\_B; 1.  
 KW Signal.

FT CHAIN 1 22 POTENTIAL.  
 FT SIGNAL 23 902 GLYCOPROTEIN B.  
 SO SEQUENCE 902 AA; 100357 MW; 6232E753A63E673 CRC64;

Query Match 100.0%; Score 36; DB 12; Length 902;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 494 SSIEFARL 501

RESULT 8  
 ID 069076 PRELIMINARY; PRT; 903 AA.

AC 069076;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE Glycoprotein B.  
 GN GB-1.  
 OS human herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10298;

RN (1)  
 RP SEQUENCE FROM N.A.

RC MEDLINE=88306231; PubMed=2841793;  
 RA Hamerschmidt W., Contrals F., Mankertz J., Paul G., Ludwig H.,  
 Buhk H.J.;

RT "Conservation of a gene cluster including glycoprotein B in bovine  
 herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";  
 RL Virology 165:386-405(1988).  
 DR EMBL: M21629; AAB19496.1; -.

DR InterPro: IPR000234; Glycoprot.B.  
 DR Pfam: PF00606; Glycoprotein\_B; 1.  
 DR Prodom: PD000693; Glycoprot\_B; 1.

SO SEQUENCE 903 AA; 100090 MW; BA2D5611906307EE CRC64;

Query Match 100.0%; Score 36; DB 12; Length 903;  
 Best Local Similarity 100.0%; Pred. No. 6.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
 DB 498 SSIEFARL 505

RESULT 9  
 ID 069526 PRELIMINARY; PRT; 904 AA.

AC 069526;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
 DE Glycoprotein B.  
 GN GB.

OS human herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.

OX NCBI\_TaxID=10298;

RN (1)  
 RP SEQUENCE FROM N.A.

RC STRAIN-ANG PATH;



RA Holland T.C., Saharkhiz-Langroodi A.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U49121; AA81805.1; -  
DR InterPro: IPR000234; Glycoprot\_B.  
DR Pfam: PF00606; Glycoprotein\_B; 1.  
DR ProDom: PD000693; Glycoprot\_B; 1.  
SQ SEQUENCE 904 AA; 100291 MW; D9C3C3B66FB1E033 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 36; DB 12; Length 904;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
|||||||  
DB 499 SSIEFARL 506

## RESULT 10

ID 090LM8 PRELIMINARY; PRT; 904 AA.  
AC 090LM8;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Glycoprotein B.  
GN UL27.  
OS human herpesvirus 1.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10298;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HSZP;  
RX MEDLINE=20227371; PubMed=10766304;  
RA Kosovsky J., Vojvodova A., Oravcova I., Kudelova M., Mattis J.,  
RA Rajcani J.;  
RT "Herpes simplex virus 1 (HSV-1) strain HSZP glycoprotein B gene :  
RT comparison of mutations among strains differing in virulence.";  
RL Virus Genes 20:27-33(2000).  
DR EMBL: AF097023; AAF04615.1; -  
DR InterPro: IPR000234; Glycoprot\_B.  
DR Pfam: PF00606; Glycoprotein\_B; 1.  
DR ProDom: PD000693; Glycoprot\_B; 1.  
SQ SEQUENCE 904 AA; 100145 MW; DF2CAD503E594D08 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 36; DB 12; Length 904;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
|||||||  
DB 499 SSIEFARL 506

## RESULT 11

ID 091WT4 PRELIMINARY; PRT; 904 AA.  
AC 091WT4;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE Glycoprotein B.  
GN UL27.  
OS human herpesvirus 1.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10298;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=+GC;  
RA Ling J.-Y., Chen T.-M., Stroop W.G.;  
RT "A viral genetic element involved in a rabbit model of herpes simplex  
RT virus-induced epileptiform seizures.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF259899; AAF70301.1; -  
DR InterPro: IPR000234; Glycoprot\_B.  
DR Pfam: PF00606; Glycoprotein\_B; 1.  
DR ProDom: PD000693; Glycoprot\_B; 1.  
SQ SEQUENCE 904 AA; 100185 MW; 9757F59FC8FEAD40 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 36; DB 12; Length 904;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
|||||||  
DB 499 SSIEFARL 506

## RESULT 12

ID 09DXE4 PRELIMINARY; PRT; 904 AA.  
AC 09DXE4;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Glycoprotein B.  
OS Herpes simplex virus (type 1 / strain KOS).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KOS;  
RX MEDLINE=21066717; PubMed=11145912;  
RA Pertel P.E., Fridberg A., Parish M.L., Spear P.G.;  
RT "Cell fusion induced by herpes simplex virus glycoproteins gB, gD, and  
RT gH-gL requires a gD receptor but not necessarily heparan sulfate.";  
RL Virology 279:313-324(2001).  
DR EMBL: AF311740; AAG34116.1; -  
DR InterPro: IPR000234; Glycoprot\_B.  
DR Pfam: PF00606; Glycoprotein\_B; 1.  
DR ProDom: PD000693; Glycoprot\_B; 1.  
SQ SEQUENCE 904 AA; 100294 MW; 21FD32F5E5F2AD48 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 36; DB 12; Length 904;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSIEFARL 8  
|||||||  
DB 499 SSIEFARL 506

## RESULT 13

ID 037453 PRELIMINARY; PRT; 904 AA.  
AC 037453;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE Glycoprotein B.  
GN UL27.  
OS Herpes simplex virus (type 2).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BBKC;  
RX MEDLINE=98314525; PubMed=9652417;  
RA Terhune S.S., Coleman K.T., Sekulovich R., Burke R.L., Spear P.G.;  
RT "Limited variability of glycoprotein gene sequences and neutralizing  
RT targets in herpes simplex virus type 2 isolates and stability on  
RT passage in cell culture.";  
RT J. Infect. Dis. 178:8-15(1998).

DR EMBL: AF021340; AAB72100.1; -  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B.1.  
 DR ProDom: PD000693; Glycoprot\_B.1.  
 SQ SEQUENCE 904 AA; 100215 MW; 2170D008154B131F CRC64;

Query Match 100.0%; Score 36; DB 12; Length 904;  
 Best Local Similarity 100.0%; Pred. NO. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSIFFARL 8  
 |||||  
 DB 496 SSIFFARL 503

RESULT 14  
 091C63 PRELIMINARY; PRT; 904 AA.  
 AC 091C63;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Glycoprotein B2.  
 GN GB2.  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID-10310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-G;  
 RA Lee H.H.;  
 RT "Sequencing and high level expression of the HSV-2 strain G  
 glycoprotein B gene using a baculovirus HcNPV."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF295528; AAK97852.1;  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B.1.  
 DR ProDom: PD000693; Glycoprot\_B.1.  
 DR SEQUENCE 904 AA; 100153 MW; 210DAA5DD2475C6 CRC64;  
 SQ

Query Match 100.0%; Score 36; DB 12; Length 904;  
 Best Local Similarity 100.0%; Pred. NO. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSIFFARL 8  
 |||||  
 DB 496 SSIFFARL 503

RESULT 15  
 089920 PRELIMINARY; PRT; 904 AA.  
 AC 089920;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Glycoprotein B precursor.  
 GN UL27.  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID-10310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WT1A;  
 RA Terhune S.S.; Spear P.G.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE OF 80-904 FROM N.A.  
 RP STRAIN-WT1A;  
 RC MEDLINE-87112925; PubMed-3027364;  
 RX Stuve L.L., Brown-Shimer S., Pachi C., Najarian R., Dina D.,

RA Burke R.L.;  
 RT "Structure and expression of the herpes simplex virus type 2  
 glycoprotein gb gene."  
 RL J. Virol. 61:326-335(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WT1A;  
 RA Spear P.G.;  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U12175; AAB60548.1; -  
 DR EMBL: U12173; AAB60546.1; -  
 DR InterPro: IPR000234; Glycoprot\_B.  
 DR Pfam: PF00606; Glycoprotein\_B.1.  
 DR ProDom: PD000693; Glycoprot\_B.1.  
 KW Signal.  
 FT SIGNAL  
 FT VARIANT 1 22 POTENTIAL.  
 FT SEQUENCE 904 AA; 100187 MW; AAB36F74FBE72139 CRC64;  
 SQ

Query Match 100.0%; Score 36; DB 12; Length 904;  
 Best Local Similarity 100.0%; Pred. NO. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSIFFARL 8  
 |||||  
 DB 496 SSIFFARL 503

Search completed: May 22, 2003, 12:09:37  
 Job time : 28 secs

GenCore version 5.1.4-P5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:16 ; Search time 31 Seconds  
(without alignments)  
34.387 Million cell updates/sec

Title: US-09-719-494-12

Perfect score: 44

Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	8	21	AAV67357
2	37	84.1	1172	23	AAU74786
3	37	84.1	1172	23	ABW72334
4	36	81.8	1199	19	AAW47206
5	36	81.8	1371	22	ABW52745
6	35	79.5	513	22	ABG07347
7	35	79.5	571	21	AAW43586
8	35	79.5	571	22	AAW5130
9	35	79.5	1172	21	AAW19677
10	35	79.5	1172	21	AAW00043

11	35	79.5	1172	23	AAU76902	Human Thrombospondin
12	35	79.5	1172	23	AAU74786	Human Thrombospondin
13	33	75.0	91	21	AAV70091	Porcine endogenous
14	33	75.0	705	22	ABG29447	Novel human diagen
15	33	75.0	1766	22	ABW58631	Drosophila melanog
16	32	72.7	8	21	AAV63360	Melanoma glycoprote
17	32	72.7	9	21	AAW03467	Human tyrosine-rel
18	32	72.7	9	22	AAU26665	Human leukocyte An
19	32	72.7	9	22	AAU26998	Human leukocyte An
20	32	72.7	10	22	AAU26707	Human leukocyte An
21	32	72.7	10	22	AAU27040	Human leukocyte An
22	32	72.7	30	23	AAU48873	Human TTP-1 segmen
23	32	72.7	30	23	AAU48938	Human TTP-1 segmen
24	32	72.7	79	23	AAU06853	Human polypeptide
25	32	72.7	134	23	ABW10041	Tyrosinase 15. Un
26	32	72.7	203	22	AAU23172	Novel human enzyme
27	32	72.7	248	23	AAU48805	Human rrp-1 consen
28	32	72.7	249	12	AAW44258	gp75 peptide and f
29	32	72.7	253	22	AAE10135	Streptomyces nours
30	32	72.7	306	17	AAW05511	HCW Toledo strain
31	32	72.7	390	22	AAW78765	Human protein SEQ
32	32	72.7	474	22	AAW79749	Human protein sequ
33	32	72.7	486	22	AAW3541	Novel human diagen
34	32	72.7	491	22	ABG07344	Novel human diagen
35	32	72.7	498	22	ABG09680	Novel human diagen
36	32	72.7	517	20	AAW42635	Murine tyrosinase-r
37	32	72.7	517	22	AAW31981	Human tyrosinase-r
38	32	72.7	517	22	AAW4679	Human tyrosinase-r
39	32	72.7	517	23	AAU76663	Mouse tyrosine-rel
40	32	72.7	517	23	AAE14345	Human protease PRI
41	32	72.7	517	23	AAU82718	Amino acid sequenc
42	32	72.7	519	18	AAW30826	The novel tyrosina
43	32	72.7	519	20	AAW42636	Human tyrosinase-r
44	32	72.7	519	20	AAW31982	Human tyrosinase-r
45	32	72.7	519	22	AAW6041	Human Ttp2 protein

#### ALIGNMENTS

RESULT 1  
AAV67357  
ID AAV67357 standard; peptide: 8 AA.  
AC AAV67357;  
XX 25-APR-2000 (first entry)  
DT  
XX Melanoma glycoprotein 75 peptide TAY used as a therapeutic antigen.  
DE  
XX  
XX  
KW Therapeutic antigen; cytotoxic T lymphocyte; CTL; CTL immune response;  
KW cellular immune response induction method; vaccine; human; tumour;  
KW melanoma glycoprotein 75.  
XX  
OS Homo sapiens.  
XX  
XX WO963945-A2.  
XX  
XX 16-DEC-1999.  
XX  
XX 11-JUN-1999; 99WO-0513146.  
XX  
XX 12-JUN-1998; 98US-0089055.  
XX 30-OCT-1998; 98US-0106339.  
XX  
XX (SLOAN) SLOAN KETTERING INST CANCER RES.  
XX NIKOLIC-ZUGLIC J, DYALL R, HOUGHTON AN;  
XX WPI, 2000-126432/11.  
XX  
XX Induction of a cellular immune response to a weakly immunogenic  
XX protein, used to target and kill tumour cells.

XX PS Claim 14; Page 24; 44pp; English.

CC This sequence represents a melanoma glycoprotein 75 peptide used as a  
 CC therapeutic antigen in the method of the invention. The invention relates  
 CC to a method for inducing a cytotoxic T lymphocyte (CTL) immune response  
 CC to non/weakly-immunogenic proteins which are expressed on tumour cells.  
 CC The method for inducing a cellular immune response to a non-immunogenic  
 CC or weakly immunogenic target peptide expressed on tumour cells of a  
 CC mammalian subject comprises administering antigen to induce a cellular  
 CC immune response to the target peptide. The antigen comprises an  
 CC immunogenic portion having a major histocompatibility complex (MHC)  
 CC binding domain which binds to the MHC and an immune recognition domain  
 CC which is recognized by T-cells. The antigen is derived from the target  
 CC peptide such that the MHC-binding portion binds to MHC with a greater  
 CC affinity than the target peptide without material alteration of the  
 CC immune recognition portion. The methods are used for inducing a cellular  
 CC immune response to a non-immunogenic or weakly immunogenic target peptide  
 CC expressed on tumour cells of a mammalian subject. The antigens and  
 CC immunogens of the invention, as well as polynucleotides encoding them,  
 CC are used in vaccine compositions against tumour cells.

SO Sequence 8 AA;

QY 1 TAYRYHLL 8  
 DB 1 TAYRYHLL 8

Query Match 100.0%; Score 44; DB 21; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
 AAU74786  
 ID AAU74786 standard; Protein: 1172 AA.

AC AAU74786;  
 DT 09-APR-2002 (first entry)  
 DE Mouse thrombospondin-2 (TSP-2).

XX Thrombospondin-2; TSP-2; cytosolic; angiogenesis; vasotropic;  
 KW vulnery; neovascularisation; cell proliferation inhibitor; cancer;  
 KW solid tumour; haemangioma; acoustic neuromas; neurofibroma; trachoma;  
 KW pyogenic granulomas; rheumatoid arthritis; ocular angiogenic disease;  
 KW retinopathy; psoriasis; macular degeneration; corneal graft rejection;  
 KW neovascular glaucoma; retrolental fibroplasia; retinoblastoma;  
 KW Osler-Webber syndrome; myocardial angiodysplasia; haemophilic joints;  
 KW plaque neovascularisation; telangiectasia; wound granulation;  
 KW mouse; apoptosis.

XX Mus sp.  
 OS WO200191781-A2.  
 PN 06-DEC-2001.  
 PD 25-MAY-2001; 2001WO-US17250.  
 PF 26-MAY-2000; 2000US-207994P.  
 PR (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA Lawler JW;  
 PI WPI; 2002-106273/14.  
 DR WPI; 2002-106273/14.  
 XX Composition useful for treatment of cancer comprises cDNA encoding  
 PT amino acids of human thrombospondin-1 or its conservative variant and a  
 PT carrier -  
 XX

PS Example 2; Fig 10; 54pp; English.

XX The invention describes a composition comprising cDNA encoding fragments  
 CC of human thrombospondin-1 (TSP-1), a type 1 repeat polypeptide and  
 CC potent inhibitor of tumour growth and angiogenesis. The composition is  
 CC useful for killing cancerous cells (preferably tumour); for reducing  
 CC volume or inhibiting growth of a tumour (inhibiting neovascularisation in  
 CC the tumour); for decreasing proliferation of tumour cells; in the  
 CC treatment of diseases and conditions associated with angiogenic activity  
 CC or misregulated growth and angiogenesis-mediated diseases such as cancer,  
 CC solid tumour, tumour metastasis, benign tumour, (e.g. haemangioma),  
 CC acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas,  
 CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g.  
 CC diabetic retinopathy, retinopathy of prematurity, macular degeneration,  
 CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasias,  
 CC telangiectasia, plaque neovascularisation, myocardial angiodysplasia,  
 CC angiodysplasia or wound granulation. The composition induces apoptosis and  
 CC inhibits neovascularisation in the tumour cells. This amino acid sequence  
 CC represents mouse thrombospondin-2 (TSP-2), on which the recombinant  
 CC protein (see AAU74796) of the invention is based.

SO Sequence 1172 AA;

QY 1 TAYRYHLL 8  
 DB 1108 TAYRYHLL 1115

Query Match 84.1%; Score 37; DB 23; Length 1172;  
 Best Local Similarity 75.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
 ABB72334  
 ID ABB72334 standard; Protein: 1172 AA.

AC ABB72334;  
 DT 04-APR-2002 (first entry)  
 DE Rat protein isolated from skin cells SEQ ID NO: 658.

XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;  
 KW developmental defect; inflammatory disease; dermatological; vulnery;  
 KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective.

XX Rattus sp.  
 OS WO200190357-A1.  
 PN 29-NOV-2001.  
 PD 24-MAY-2001; 2001WO-NZ00099.  
 PF 24-MAY-2000; 2000US-206650P.  
 PR 25-JUL-2000; 2000US-221232P.  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PI Watson JD, Strachan L, Sleeman M, Onrust R, Morrison JG, Kumble KD;  
 DR WPI; 2002-122020/16.  
 DR N-PSDB; ABL35019.  
 XX New polynucleotides and polypeptides encoded by the polynucleotides  
 PT isolated from skin cells, useful for treating skin wounds, cancers,  
 PT growth and developmental defects, inflammatory diseases, or for  
 PT modulating immune responses -  
 XX  
 PS Claim 4; Page 413-416; 466pp; English.  
 CC The present invention provides the protein and coding sequences of cDNAs

CC isolated from human, murine and rat skin cell libraries. The sequences  
 CC can be used in the development of therapeutic agents useful in the  
 CC treatment of skin diseases, including skin wounds, cancer, growth  
 CC defects, developmental defects and inflammatory diseases. The proteins  
 CC have important roles in the induction of hair growth, cell proliferation  
 CC and cell-cell interaction. In maintaining tissue integrity, in wound  
 CC healing and in modulating immune responses. The present sequence is a  
 CC polypeptide of the invention.

XX  
 SQ Sequence 1172 AA;

Query Match 84.1%; Score 37; DB 23; Length 1172;  
 Best Local Similarity 75.0%; Pred. No. 1e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAYRYHL 8  
 |||||:  
 Db 1108 TAYRYHL 1115

RESULT 4  
 AAM47206  
 ID AAM47206 standard; Protein; 1199 AA.

XX  
 AC AAM47206;

XX  
 DT 21-JUL-1998 (first entry)

XX  
 DE Bos taurus tubulin-folding cofactor D.

XX  
 KM Tubulin folding; cofactor: alpha-tubulin; beta-tubulin; unfolded;  
 KM folded; treatment; hyper-proliferative diseases; cancer; gout.

XX  
 OS Bos taurus.

XX  
 PN WO9804587-A1.

XX  
 PD 05-FEB-1998.

XX  
 PF 25-JUL-1997; 97WO-US14076.

XX  
 PR 25-JUL-1996; 96US-0023089.

XX  
 PA (UTNY ) UNIV NEW YORK STATE.

XX  
 PI Cowan NJ;

XX  
 DR WPI; 1998-130618/12.

XX  
 DR N-PSDB; AAV17086.

PT New isolated cofactor(s) for tubulin folding - are useful as targets  
 PT for identifying agents which interfere with folding in the treatment  
 PT of hyper-proliferative diseases such as cancer

XX  
 PS Claim 3; Pages 48-52; 87pp; English.

XX  
 CC The sequence is that of bovine tubulin-folding cofactor D.  
 CC it may be useful as a target for interfering with the  
 CC production of productively folded alpha- and beta-tubulins.  
 CC Since tubulin function is essential for cell division and  
 CC proliferation, agents which interfere with tubulin function  
 CC can serve as useful antiproliferative compounds. Such interfering  
 CC agents have potential utility as agents for the treatment of  
 CC hyperproliferative diseases such as cancer and the treatment  
 CC of gout.

XX  
 SQ Sequence 1199 AA;

Query Match 81.8%; Score 36; DB 19; Length 1199;  
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AYRYHL 8

Db 985 AYRYHL 991  
 |||||:

RESULT 5  
 ABB52745  
 ID ABB52745 standard; Protein; 1371 AA.

XX  
 AC ABB52745;

XX  
 DT 11-FEB-2002 (first entry)

XX  
 DE Escherichia coli polypeptide SEQ ID NO 891.

XX  
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
 KW systemic infection; non-diarrhoeal infection; septicemia;  
 KW pyelonephritis; antibiotic resistance.

XX  
 OS Escherichia coli.

XX  
 PN WO200166572-A2.

XX  
 PD 13-SEP-2001.

XX  
 PF 12-MAR-2001; 2001WO-EP03445.

XX  
 PR 10-MAR-2000; 2000FR-0003145.

XX  
 PR 02-FEB-2001; 2001FR-0001449.

XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX  
 PI Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;

XX  
 DR WPI; 2001-550253/61.

PT A library of DNA fragments of Escherichia coli strains for the  
 PT phylogenic determination of a given strain comprises polynucleotides of  
 PT nature B2/D+ A- -  
 PS Example 6; Fig 6; 646pp; English.

XX  
 CC The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533).  
 CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature  
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as  
 CC septicaemia, pyelonephritis and meningitis this is particularly  
 CC advantageous as bacterial resistance is increasing with the more  
 CC frequent use of broad spectrum antibiotics.

XX  
 SQ Sequence 1371 AA;

Query Match 81.8%; Score 36; DB 22; Length 1371;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TAYRYHL 7  
 |||||:  
 Db 1226 TGYRYHL 1232

RESULT 6

ABG07347  
 ID ABG07347 standard; Protein; 513 AA.

XX  
 AC ABG07347;

XX

DT 13-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #7338.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HSE-) HSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB; AAS71534.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX Claim 20; SEQ ID NO 37706; 103bp; English.  
 PS The invention relates to isolated polynucleotide (I) and  
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 513 AA;  
 SQ  
 Query Match 79.5%; Score 35; DB 22; Length 513;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 YRYHLL 8  
 Db 493 YRYHLL 498  
 RESULT 7  
 AAB43586  
 ID AAB43586 standard; Protein: 571 AA.  
 XX AAB43586;  
 XX 08-FEB-2001 (first entry)  
 DT

DE Human cancer associated protein sequence SEQ ID NO:1031.  
 KW Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vunerary; immunomodulator;  
 KW antidiabetic; antiaesthetic; antineumatic; antiallergic; antiviral;  
 KW antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;  
 KW vasotropic; antipsoriatic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.  
 XX Homo sapiens.  
 OS WO200055350-A1.  
 XX 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US05882.  
 XX 12-MAR-1999; 99US-0124270.  
 XX (HDMA-) HDMA GENOME SCI INC.  
 PA Rosen CA, Ruben SM;  
 PI WPI: 2000-587533/55.  
 DR N-PSDB; AAC77795.  
 XX Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer  
 PT Claim 11; Page 1616-1619; 2352bp; English.  
 PS AAC77607 to AAC78448 encode the human cancer associated proteins given  
 XX in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vunerary; immunomodulator;  
 CC antidiabetic; antiaesthetic; antineumatic; antiallergic; antiviral;  
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; cardiant;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC nocotropic; vasotropic; antipsoriatic and angiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.  
 XX Sequence 571 AA;  
 SQ  
 Query Match 79.5%; Score 35; DB 21; Length 571;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAYRYHLL 7  
 Db 507 TAYRYHLL 513  
 RESULT 8  
 AAG75130  
 ID AAG75130 standard; Protein: 571 AA.  
 XX AAG75130;  
 AC

```

XX 03-SEP-2001 (first entry)
DT
DE Human colon cancer antigen protein SEQ ID NO:5894.
KM Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 6.
XX
OS Homo sapiens.
PN WO200122920-A2.
PD
PF 05-APR-2001.
PR 28-SEP-2000; 2000WO-US26524.
PS 29-SEP-1999; 99US-0157137.
PT 03-NOV-1999; 99US-0163280.
PI (HUMA-) HUMAN GENOME SCI INC.
PA
PB Ruben SM, Barash SC, Birse CE, Rosen CA;
PC WPI: 2001-235357/24.
PD N-PDB: AAH34535.
PE
PF Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PX Claim 11; Page 7409-7411; 9803pp; English.
PS
PP AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
CX
SQ Sequence 571 AA;
Query Match 79.5%; Score 35; DB 22; Length 571;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 TAYRYHL 7
| | | : | |
| | | : | |
Db 507 TAYRYHL 513
RESULT 9
ID AAB19677 standard; Protein: 1172 AA.
XX AAB19677;
AC AAB19677;
DT 05-FEB-2001 (first entry)
DE Human thrombospondin-2.
DX Thrombospondin-2; TSP-2; human; angiogenesis; cell proliferation;
TX
```

KW melanoma; tumour; cancer; squamous cell carcinoma; antiangiogenic;  
 KW prostate cancer; psoriasis; rosacea dermatosis; antitumour;  
 KW therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Region 382..429  
 FT Region /note="type 1 repeat"  
 FT Region 384..390  
 FT Region /note="WSPFAM sequence involved in antiangiogenic  
 FT Region activity"  
 FT Region 438..490  
 FT Region /note="type 1 repeat"  
 FT Region 495..547  
 FT Region /note="type 1 repeat"  
 PN  
 XX MO200057899-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 24-MAR-2000; 2000WO-US07835.  
 PE  
 XX 31-MAR-1999; 99US-0127221.  
 PR  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA  
 XX  
 PI Detmar M, Strelt M;  
 PI  
 XX WPI: 2000-656131/63.  
 DR  
 XX N-PSDB: AAA88669.  
 DX  
 PT Treating a disorder characterized by unwanted cell proliferation e.g.  
 PT precancerous, cancerous or neoplastic cells or presence of tumor  
 PT preferably of skin or prostate, comprises increasing thrombospondin-2  
 PT activity -  
 PS  
 XX  
 PS Disclosure; Flg 2; 73pp; English.  
 XX  
 XX The present sequence is that of human thrombospondin-2 (TSP-2).  
 CC The invention is based on the discovery that overexpression of  
 CC TSP-2 decreases tumor size in vivo, and features methods for  
 CC modulating unwanted angiogenesis and tumor growth. Treatment of  
 CC unwanted cell proliferation or angiogenesis involves increasing  
 CC TSP-2 activity. This is achieved by administering an agent which  
 CC increases TSP-2 activity, especially a TSP-2 polypeptide, a TSP-2  
 CC derived polypeptide or retro-inverso peptide, a nucleic acid  
 CC encoding TSP-2, an agonist of TSP-2, or an agent that increases  
 CC TSP-2 gene expression. The TSP-2 polypeptide may include at least  
 CC one type 1 repeat such as the WSPFAM peptide (see AA019683),  
 CC which is involved in the antiangiogenic activity of TSP-2. The  
 CC method is used to treat a disorder characterized by pre-cancerous,  
 CC cancerous or neoplastic cells, or the presence of a tumor, or a  
 CC disorder that affects epithelial tissues resulting in unwanted  
 CC skin cell proliferation. Such disorders include malignant  
 CC melanoma, prostate cancer, squamous cell carcinoma, aged skin,  
 CC rosacea dermatosis, psoriasis, and skin damage caused by  
 CC photoradiation (all claimed). Evaluating the presence of TSP-2  
 CC nucleic acid or protein is useful for diagnosing a subject at risk  
 CC of unwanted cell proliferation or angiogenesis. Methods are also  
 CC provided of identifying compounds that modulate TSP-2 activity.  
 CC  
 XX  
 XX Sequence 1172 AA;  
 SQ  
 Query Match: 79.5%; Score 35; DB 21; Length 1172;  
 Best Local Similarity 85.7%; Pred. NO. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAYRYHL 7  
 |||||  
 DB 1108 TAYRYHL 1114





XX DE Human thrombospondin-2 (TSP-2).  
 XX DE  
 XX Thrombospondin-2; TSP-2; cytostatic; angiogenesis; vasotropic;  
 KM vulnerable; neovascularisation; cell proliferation inhibitor; cancer;  
 KM solid tumour; haemangioma; acoustic neuromas; neurofibroma; trachoma;  
 KM pyogenic granulomas; rheumatoid arthritis; ocular angiogenic disease;  
 KM retinopathy; psoriasis; macular degeneration; corneal graft rejection;  
 KM neovascular glaucoma; retrolental fibroplasia; rebeosis; angiodioma;  
 KM Osler-Weber syndrome; myocardial angiogenesis; haemophilic joints;  
 KM plaque neovascularisation; telangiectasia; wound granulation; apoptosis.  
 XX KM  
 OS Homo sapiens.  
 PN WO200191781-A2.  
 PD 06-DEC-2001.  
 XX PD  
 XX 25-MAY-2001; 2001WO-US17250.  
 PF 25-MAY-2001; 2001WO-US17250.  
 XX PF  
 XX 26-MAY-2000; 2000US-207994P.  
 PR 26-MAY-2000; 2000US-207994P.  
 XX PR  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX PA  
 XX Lawler JW;  
 PI Lawler JW;  
 XX PI  
 XX WPI; 2002-106273/14.  
 DR WPI; 2002-106273/14.  
 XX DR  
 PT Composition useful for treatment of cancer comprises cDNA encoding  
 PT amino acids of human thrombospondin-1 or its conservative variant and a  
 PT carrier -  
 PT carrier -  
 XX PT  
 XX Disclosure; Fig 11; 54pp; English.  
 PS Disclosure; Fig 11; 54pp; English.  
 XX PS  
 XX The invention describes a composition comprising cDNA encoding fragments  
 CC of human thrombospondin-1 (TSP-1), a type 1 repeat polypeptide and is  
 CC potent inhibitor of tumour growth and angiogenesis. The composition is  
 CC useful for killing cancerous cells (preferably tumour); for reducing  
 CC volume or inhibiting growth of a tumour (inhibiting neovascularisation in  
 CC the tumour); for decreasing proliferation of tumour cells; in the  
 CC treatment of diseases and conditions associated with angiogenic activity  
 CC or misregulated growth and angiogenesis-mediated diseases such as cancer,  
 CC solid tumour, tumour metastasis, benign tumour, (e.g. haemangioma,  
 CC acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas),  
 CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g.  
 CC diabetic retinopathy, retinopathy of prematurity, macular degeneration,  
 CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasias,  
 CC rebeosis), Osler-Weber syndrome, myocardial angiogenesis,  
 CC telangiectasia, plaque neovascularisation, haemophilic joints,  
 CC angiodioma or wound granulation. The composition induces apoptosis and  
 CC inhibits neovascularisation in the tumour cells. This amino acid sequence  
 CC represents human thrombospondin-2 (TSP-2), on which the recombinant  
 CC protein (shown in AAU74789) of the invention is based.  
 CC CC  
 XX CC  
 SQ Sequence 1172 AA;  
 Query Match 79.5%; Score 35; DB 23; Length 1172;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAYRYHL 7  
 | | | | | | |  
 Db 1108 TAYRYMHL 1114

DE Porcine endogenous retrovirus-D env polypeptide.  
 XX DE  
 XX Porcine endogenous retrovirus; PERV-D; virucide; prevention; vaccine;  
 KM pig; diagnosis; infection; xenotransplantation; antibody; env region.  
 XX KM  
 OS Sus scrofa.  
 XX OS  
 XX WO200011187-A1.  
 PN WO200011187-A1.  
 XX PN  
 PD 02-MAR-2000.  
 XX PD  
 XX 18-AUG-1999; 99WO-US19053.  
 PF 18-AUG-1999; 99WO-US19053.  
 XX PF  
 XX 18-AUG-1998; 98US-0097015.  
 PR 18-AUG-1998; 98US-0097015.  
 XX PR  
 XX (BIOT-) BIO TRANSPLANT INC.  
 PA (BIOT-) BIO TRANSPLANT INC.  
 XX PA  
 XX Banerjee PT, Patience C, Andersson GK;  
 PI Banerjee PT, Patience C, Andersson GK;  
 XX PI  
 XX WPI; 2000-224704/19.  
 DR WPI; 2000-224704/19.  
 XX DR  
 XX N-PSDB; AA251065.  
 DR N-PSDB; AA251065.  
 XX DR  
 XX Porcine retroviral PERV-D polypeptides for diagnosing porcine  
 PT retroviral infections in humans after xenotransplantation -  
 PT retroviral infections in humans after xenotransplantation -  
 XX PT  
 XX Claim 9; Fig 2; 119pp; English.  
 PS Claim 9; Fig 2; 119pp; English.  
 XX PS  
 XX The present amino acid sequence is the porcine endogenous retrovirus  
 CC env polypeptide-D (PERV-D). It is isolated from the porcine kidney, PK15  
 CC cell line (ATCC No. CCL-33). PERV-D polynucleotide has 79% homology to a  
 CC portion of PERV-C and has virucidal activity. PERV-D sequence is useful  
 CC for prevention or diagnosis of infection of human tissues by porcine  
 CC retroviruses after xenotransplantation procedures. PERV-D polypeptide  
 CC may be used to produce specific antibodies, that can be administered as  
 CC vaccines to create passive immunity. The DNA can be used as a  
 CC hybridisation probe or primer for isolation purposes.  
 CC CC  
 XX CC  
 SQ Sequence 91 AA;  
 Query Match 75.0%; Score 33; DB 21; Length 91;  
 Best Local Similarity 83.3%; Pred. No. 40;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAYRYH 6  
 | | | | |  
 Db 23 TSYRYH 28

RESULT 14  
 ABG29447  
 ID ABG29447 standard; Protein; 705 AA.  
 XX ID  
 XX ABG29447;  
 AC ABG29447;  
 XX AC  
 XX 18-FEB-2002 (first entry)  
 DT 18-FEB-2002 (first entry)  
 XX DT  
 XX Novel human diagnostic protein #29438.  
 DE Novel human diagnostic protein #29438.  
 XX DE  
 KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX KM  
 OS Homo sapiens.  
 XX OS  
 XX WO200175067-A2.  
 PN WO200175067-A2.  
 XX PN  
 PD 11-OCT-2001.  
 XX PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX PF  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 31-MAR-2000; 2000US-0540217.  
 XX PR  
 XX 23-AUG-2000; 2000US-0649167.  
 XX 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.

```

XX  Drmanac RT, Liu C, Tang YT;
XX  WPI: 2001-639362/73.
XX  N-PSDB: AAS93634.
XX  New isolated polynucleotide and encoded polypeptides, useful in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits and to assess
XX  biodiversity -
XX  Claim 20; SEQ ID NO 59806; 103pp; English.
XX  The invention relates to isolated polynucleotide (I) and
XX  polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX  and gene mapping, and in recombinant production of (II). The
XX  polynucleotides are also used in diagnostics as expressed sequence tags
XX  for identifying expressed genes. (I) is useful in gene therapy techniques
XX  to restore normal activity of (II) or to treat disease states involving
XX  (II). (II) is useful for generating antibodies against it, detecting or
XX  quantitating a polypeptide in tissue, as molecular weight markers and as
XX  a food supplement. (II) and its binding partners are useful in medical
XX  imaging of sites expressing (II). (I) and (II) are useful for treating
XX  disorders involving aberrant protein expression or biological activity.
XX  The polypeptide and polynucleotide sequences have applications in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. AAG00010-ABG30377 represent novel human
XX  diagnostic amino acid sequences of the invention.
XX  Note: The sequence data for this patent did not appear in the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX  SQ Sequence 705 AA;
XX
XX  Query Match 75.0%; Score 33; DB 22; Length 705;
XX  Best Local Similarity 71.4%; Pred. No. 3.6e+02;
XX  Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY 1 TAYRYHL 7
XX  1::11111
XX  542 TSFRYHL 548
XX
XX  RESULT 15
XX  ABB58631
XX  ID ABB58631 standard; Protein; 1766 AA.
XX  AC ABB58631;
XX  DT 26-MAR-2002 (first entry)
XX  DE Drosophila melanogaster polypeptide SEQ ID NO 2685.
XX  DE Drosophila; developmental biology; cell signalling; insecticide;
XX  KW pharmaceutical.
XX  OS Drosophila melanogaster.
XX  PN WO200171042-A2.
XX  PD 27-SEP-2001.
XX  PF 23-MAR-2001; 2001WO-US09231.
XX  PR 23-MAR-2000; 2000US-191637P.
XX  PR 11-JUL-2000; 2000US-0614150.
XX  PA (PEKE ) PE CORP NY.
XX  PI Venter JC, Adams M, Li PWD, Myers EW;

```

```

XX  WPI: 2001-656860/75.
XX  N-PSDB: ABL02734.
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
XX  genes from Drosophila and for elucidating cell signalling and cell-cell
XX  interactions -
XX  Disclosure; SEQ ID NO 2685; 21pp + Sequence Listing; English.
XX  The invention relates to an isolated nucleic acid detection reagent
XX  capable of detecting 1000 or more genes from Drosophila. The invention is
XX  useful in developmental biology and in elucidating cell signalling and
XX  cell-cell interactions in higher eukaryotes for the development of
XX  insecticides, therapeutics and pharmaceutical drugs. The invention
XX  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX  sequences (ABB57737-ABB72072).
XX  The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX  SQ Sequence 1766 AA;
XX
XX  Query Match 75.0%; Score 33; DB 22; Length 1766;
XX  Best Local Similarity 85.7%; Pred. No. 9.7e+02;
XX  Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX  QY 2 AYRYHLL 8
XX  111111
XX  1399 AYRYHLL 1405
XX
XX  Search completed: May 22, 2003, 12:08:19
XX  Job time : 32 secs

```

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:06:12 ; Search time 14 seconds  
(without alignments)  
54.934 Million cell updates/sec

Title: US-09-719-494-12

Perfect score: 44  
Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	37	84.1	1172	2 A42587	thrombospondin 2 p
2	37	84.1	1178	1 A39804	thrombospondin pre
3	36	81.8	145	2 F75337	transposase - Dein
4	36	81.8	1199	2 T18522	tubulin-folding co
5	35	79.5	323	2 F98303	probable transcrip
6	35	79.5	323	2 AG2979	transcription regu
7	35	79.5	944	2 T47246	chitin synthase (E
8	35	79.5	1172	1 TSHMP2	thrombospondin 2 p
9	34	77.3	162	2 A70489	conserved hypochet
10	33	75.0	72	2 T17937	hypothetical prote
11	33	75.0	173	2 D72586	hypothetical prote
12	33	75.0	321	2 G70415	nucleotide sugar e
13	33	75.0	374	2 H81399	2-oxoglutarate-fer
14	32	72.7	150	2 S27613	hypothetical prote
15	32	72.7	172	2 AC2548	hypothetical prote
16	32	72.7	214	2 S44706	opacly protein op
17	32	72.7	234	2 S36329	opacly protein op
18	32	72.7	234	2 S36342	opacly protein op
19	32	72.7	234	2 S36341	opacly protein op
20	32	72.7	235	2 S44707	opacly protein op
21	32	72.7	248	2 P10038	opacly protein D
22	32	72.7	254	2 S20043	opacly protein B
23	32	72.7	257	2 S16614	opacly protein op
24	32	72.7	258	2 S08514	opacly protein-re
25	32	72.7	268	1 KONR2C	opacly protein P.
26	32	72.7	283	2 S72343	opacly protein op
27	32	72.7	294	2 H83490	hypothetical prote
28	32	72.7	338	2 S16613	opacly protein op
29	32	72.7	352	2 I51282	TRP-1 - axolotl (f

30	32	72.7	366	2 T12548	hypothetical prote
31	32	72.7	414	2 S18962	FBF15 protein - St
32	32	72.7	517	1 S19243	tyrosinase-related
33	32	72.7	519	1 YRHUR2	dopachrome isomera
34	32	72.7	522	2 I51245	tyrosinase related
35	32	72.7	537	1 YRHUB6	tyrosinase-related
36	32	72.7	537	1 YRHUB6	tyrosinase-related
37	32	72.7	550	2 A10807	probable decarboxy
38	32	72.7	679	2 T00636	hypothetical prote
39	32	72.7	1050	2 T43482	hypothetical prote
40	32	72.7	3951	2 VFTMB1	fl protein - avian
41	31	70.5	72	1 S31028	gene 83 protein -
42	31	70.5	121	1 PC4024	phospholipase A2 h
43	31	70.5	129	1 C87219	hypothetical prote
44	31	70.5	141	2 A70556	probable mutator M
45	31	70.5	190	2 S18680	DNA-Invertase - Es

#### ALIGNMENTS

##### RESULT 1

A42587  
thrombospondin 2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Mar-1993 #sequence\_rev150n 18-Nov-1994 #text\_change 20-Aug-1999  
C:Accession: A42587; A39851  
R:Laberty, C.D., O'Rourke, K., Wolf, F.W., Katz, R., Seldin, M.F., Dixit, V.M.  
J. Biol. Chem. 267, 3274-3281, 1992  
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during  
A:Reference number: A42587; MIMD:92147663; PMID:1371115  
A:Accession: A42587  
A>Status: Preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 11172 <LAB>  
A:Cross-references: GB:U07803; GB:M87275; NID:9340421; PIDN:AAA53064.1; PID:9567241  
A>Note: sequence extracted from NCBI backbone (NCBI:81502)  
R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.  
J. Biol. Chem. 266, 12821-12824, 1991  
A:Title: A second, expressed thrombospondin gene (Tbbs2) exists in the mouse genome  
A:Reference number: A39851; MIMD:91302287; PMID:1712771  
A:Accession: A39851  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-873 <BOR>  
A:Cross-references: GB:M64866; NID:9201994; PIDN:AAA40432.1; PID:9201995  
C:Superfamily: Thrombospondin 1; EGF homology; Thrombospondin type 1 repeat homology  
C:Keywords: calcium binding; glycoprotein  
F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>  
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>  
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>  
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>  
F:553-588/Domain: EGF homology <EGF1>  
F:652-691/Domain: EGF homology <EGF>  
Query Match 84.1%; Score 37; DB 2; Length 1172;  
Best Local Similarity 75.0%; Pred. No. 28;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TAYRYHLL 8  
DB 1108 TAYRYHLL 1115  
RESULT 2  
A39804  
thrombospondin precursor - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Sep-1999 #sequence\_rev150n 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A39804  
R:Lawler, J.; Duquette, M.; Ferro, P.  
J. Biol. Chem. 266, 8039-8043, 1991  
A:Title: Cloning and sequencing of chicken thrombospondin.

A:Reference number: A39804; MUID:91217026; PMID:2022631  
 A:Accession: A39804  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1178 <LMW>  
 A:Cross-references: GB:M60853; NID:9212763; PIDN:AAA51437.1; PID:9212764  
 C:Superfamily: Thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; VWC  
 F:325-383/Domain: von Willebrand factor type C repeat homology <THR1>  
 F:386-437/Domain: thrombospondin type 1 repeat homology <THR2>  
 F:442-498/Domain: thrombospondin type 1 repeat homology <THR3>  
 F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>  
 F:658-697/Domain: EGF homology <EGF>

Query Match 84.1%; Score 37; DB 1; Length 1178;  
 Best Local Similarity 75.0%; Pred. No. 28;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
 DB 1114 TAYRYHLL 1121

RESULT 3  
 F75337  
 transposase - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: F75337  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mc  
 Science 286, 1571-1577, 1999  
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A5250; MUID:20036896; PMID:1056266  
 A:Accession: F75337  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-145 <MWI>  
 A:Cross-references: GB:AE002032; GB:AE00513; NID:96459715; PIDN:AAI1477.1; PID:9645970  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1927  
 A:Map position: 1  
 C:Superfamily: conserved hypothetical protein 4 (insertion sequence ISH1.8)

Query Match 81.8%; Score 36; DB 2; Length 145;  
 Best Local Similarity 75.0%; Pred. No. 5;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
 DB 22 TAYRYHLL 29

RESULT 4  
 T18522  
 tubulin-folding cofactor D - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 28-Jul-2000  
 C:Accession: T18522  
 R:Tian, G.; Huang, Y.; Rommelaere, H.; Vandekerckhove, J.; Ampe, C.; Cowan, N.J.  
 Cell 86, 287-296, 1996  
 A>Title: Pathway leading to correctly folded beta-tubulin.  
 A:Reference number: Z18945; MUID:96319731; PMID:8706133  
 A:Accession: T18522  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1199 <TIA>  
 A:Cross-references: EMBL:U61233; NID:91465769; PID:91465770; PIDN:AAI17537.1  
 C:Function:  
 A:Description: cofactors A and D are involved in capturing and stabilizing beta-tubulin  
 C:Superfamily: Arabidopsis thaliana beta-tubulin cofactor-like protein

Query Match 81.8%; Score 36; DB 2; Length 1199;  
 Best Local Similarity 85.7%; Pred. No. 46;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 AYRYHLL 8  
 DB 985 AYRYHLL 991

RESULT 5  
 F98303  
 probable transcription regulator Pail84 [imported] - Agrobacterium tumefaciens (stra  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
 C:Accession: F98303  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold  
 A.; Liu, F.; Woliam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,  
 Science 294, 2323-2328, 2001  
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
 A:Reference number: A97359; PMID:11743194  
 A:Accession: F98303  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-323 <KUR>  
 A:Cross-references: GB:AE007870; PIDN:AAK89952.1; PID:91515911; GSPDB:GN00170  
 C:Genetics:  
 A:Gene: AGR\_L\_2774  
 A:Map position: linear chromosome  
 C:Superfamily: regulatory protein ampr

Query Match 79.5%; Score 35; DB 2; Length 323;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
 DB 287 TAYRYHLL 294

RESULT 6  
 AG2979  
 transcription regulator, lysR family Atu3440 [imported] - Agrobacterium tumefaciens  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
 C:Accession: AG2979  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wo  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McC  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka  
 ster, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AG2979  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-323 <KUR>  
 A:Cross-references: GB:AE006889; PIDN:AAI4253.1; PID:917741837; GSPDB:GN00187  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu3440  
 A:Map position: linear chromosome  
 C:Superfamily: regulatory protein ampr

Query Match 79.5%; Score 35; DB 2; Length 323;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
 DB 287 TAYRYHLL 294

## RESULT 7

T47246  
chitin synthase (EC 2.4.1.16) 2 (Imported) - Neurospora crassa  
C:Species: Neurospora crassa  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000  
C:Accession: T47246  
R:Beth Din, A.; Yarden, O.  
Microbiology 140, 2189-2197, 1994  
A:Title: The Neurospora crassa chs-2 gene encodes a non-essential chitin synthase.  
A:Reference number: 224422; M0ID:95039879; PMID:7952169  
A:Accession: T47246  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-944 <BET>  
A:Cross-references: EMBL:X77782; NID:g558230; PIDN:CAA54816.1; PID:g558604  
C:Genetics:  
A:Gene: chs-2  
A:Map position: IVR  
A:Introns: 542/1  
C:Superfamily: chitin synthase chsA  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 79.5%; Score 35; DB 2; Length 944;  
Best Local Similarity 75.0%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
|||||  
Db 468 SAYRYHAL 475

## RESULT 8

TSHUP2  
thrombospondin 2 precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 19-May-1995 #sequence\_revision 03-Aug-1995 #text\_change 13-Aug-1999  
C:Accession: A47379; A42173  
R:Labell, T.L.; Byers, P.H.  
Genomics 17, 225-229, 1993  
A:Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: pote  
A:Reference number: A47379; M0ID:94010892; PMID:8406456  
A:Accession: A47379  
A:Molecule type: mRNA  
A:Residues: 1-1172 <LAB>  
A:Cross-references: GB:L12350; NID:g307505; PIDN:AAA03703.1; PID:g307506  
R:Labell, T.L.; Mlewicz, D.J.; Distchech, C.M.; Byers, P.H.  
Genomics 12, 421-429, 1992  
A:Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expression c  
A:Reference number: A42173; M0ID:92217961; PMID:1559694  
A:Accession: A42173  
A:Molecule type: mRNA  
A:Residues: 560-1172 <LA2>  
A:Cross-references: GB:M81339  
A:Experimental source: fibroblast  
A:Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:95096)  
C:Genetics:  
A:Gene: GDB:THB2; TSP2  
A:Cross-references: GDB:128789; OMIM:188061  
A:Map position: 6q27-6q27  
C:Complex: homotrimer, disulfide linked  
C:Function:  
A:Description: participates in cell migration and adhesion, and in platelet aggregation  
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vc  
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; trimer  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-1172/Product: thrombospondin 2 #status predicted <MAT>  
F:319-377/Domain: von Willebrand factor type C repeat homology <WVC>  
F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>  
F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>  
F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>  
F:553-588/Domain: EGF homology <EGF1>  
F:652-691/Domain: EGF homology <EGF>  
F:928-930/Region: cell attachment (R-G-D) motif

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (asn) (covalent) #status  
F:167-226/Disulfide bonds: #status predicted  
F:266,270/Disulfide bonds: interchain #status predicted  
F:612/Modified site: erythro-beta-hydroxyasparagine (asn) #status predicted

Query Match 79.5%; Score 35; DB 1; Length 1172;  
Best Local Similarity 85.7%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYHLL 7  
|||||  
Db 1108 TAYRYHLL 1114

## RESULT 9

A70489  
conserved hypothetical protein aq\_2197 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 11-Jan-2000  
C:Accession: A70489  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; M0ID:9819666; PMID:9537320  
A:Accession: A70489  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-162 <AOE>  
A:Cross-references: GB:AE000777; NID:g2984377; PIDN:AMC07909.1; PID:g2984391; GB:AE  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: aq\_2197  
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repe  
F:11-43/Domain: tetratricopeptide repeat homology #status atypical <TR1>  
F:44-77/Domain: tetratricopeptide repeat homology <TR2>  
F:78-111/Domain: tetratricopeptide repeat homology <TR3>  
F:114-147/Domain: tetratricopeptide repeat homology <TR4>

Query Match 77.3%; Score 34; DB 2; Length 162;  
Best Local Similarity 71.4%; Pred. No. 14;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 AYRYHLL 8  
|||||  
Db 18 AYRYHLL 24

## RESULT 10

T17937  
hypothetical protein a434L - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17937  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: 218806  
A:Accession: T17937  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-72 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96602.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Note: a434L

Query Match 75.0%; Score 33; DB 2; Length 72;  
Best Local Similarity 71.4%; Pred. No. 9.7;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 AYRYHLL 8  
|||||  
Db 55 AYRYHLL 61

RESULT 11  
D72586  
hypothetical protein APE1161 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 08-Sep-2000  
C:Accession: D72586  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyperthermophilic Crenarchaeon, Aeropyrum A:Reference number: A72450; NCBI:9310339; PMID:10382966  
A:Accession: D72586  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-173 <KAW>  
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80146.1; PID:d1043932; PID:9510 A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1161  
C:Superfamily: Aeropyrum pernix hypothetical protein APE1161

Query Match  
Best Local Similarity 75.0%; Score 33; DB 2; Length 173;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 YRYHLL 8  
|||||  
DB 125 YRYHLL 130

RESULT 12  
G70415  
nucleotide sugar epimerase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999  
C:Accession: G70415  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; NCBI:9819666; PMID:9537320  
A:Accession: G70415  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-321 <AOF>  
A:Cross-references: GB:AE000735; NID:92983749; PIDN:AAC07310.1; PID:92983750; GB:AE00065 A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: use  
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology F/3-316/Domain: UDPglucose 4-epimerase homology <NDP>

Query Match  
Best Local Similarity 75.0%; Score 33; DB 2; Length 321;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
|||||  
DB 161 TAYRYHLL 168

RESULT 13  
H81399  
2-oxoglutarate-ferredoxin oxidoreductase (EC 1.2.7.-) alpha chain Cj0536 [similarity] - C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: H81399  
R:Parthill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chittif C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: AB1250; NCBI:20150912; PMID:10688204  
A:Accession: H81399  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-374 <PAR>  
A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAB75172.1; PID:969 A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: coxA; Cj0536  
C:Superfamily: Helicobacter pylori 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid fe C:Keywords: oxidoreductase

Query Match  
Best Local Similarity 75.0%; Score 33; DB 2; Length 374;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 7  
|||||  
DB 220 TGYRYHV 226

RESULT 14  
S27613  
hypothetical protein 1 - Streptomyces aureofaciens (ATCC 10762) plasmid pop2621 (fira C:Species: Streptomyces aureofaciens  
A:Variety: ATCC 10762  
C:Date: 22-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 15-Jun-2001  
C:Accession: S27613  
R:Pelletier, I.; Pelfer, O.; Altenbuchner, J.; van Pee, K.H.  
submitted to the EMBL Data Library, February 1992  
A:Description: Molecular cloning and sequencing of a non haem bromoperoxidase from S A:Reference number: S27613  
A:Accession: S27613  
A:Molecule type: DNA  
A:Residues: 1-150 <PEL>  
A:Cross-references: EMBL:M84990; NID:9150455; PIDN:AAB84314.1; PID:9150456 C:Genetics:  
A:Gene: plasmid pop2621

Query Match  
Best Local Similarity 72.7%; Score 32; DB 2; Length 150;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHLL 8  
|||||  
DB 64 TAYRYHLL 71

RESULT 15  
AC2548  
hypothetical protein alr7670 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc71 C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC2548  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Matsumoto, A.; Iritg Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabat DNA Res. 6, 205-213, 2001  
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium A:Reference number: AB1807; NCBI:21595285; PMID:11759840  
A:Accession: AC2548  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <KUR>  
A:Cross-references: GB:AP003602; PIDN:BA877313.1; PID:917134755; GSPDB:GN00181 A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr7670  
A:Genome: plasmid

Query Match  
Best Local Similarity 72.7%; Score 32; DB 2; Length 172;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Thu May 22 14:11:17 2003

us-09-719-494-12.rpt

Page 5

OY 2 AYRHL 7  
| | | | :  
Db 136 AYRHHV 141

Search completed: May 22, 2003, 12:10:13  
Job time : 16 secs

---

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:01:57 ; Search time 7.5 Seconds

(without alignments)  
44.241 Million cell updates/sec

Title: US-09-719-494-12

Perfect score: 44

Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	37	84.1	1172 1	TSP2_MOUSE
2	37	84.1	1178 1	TSP2_CHICK
3	35	79.5	944 1	CHS2_NEUCR
4	35	79.5	988 1	CHS1_EXODE
5	35	79.5	1043 1	CHS2_PABBR
6	35	79.5	1170 1	TSP1_BOVIN
7	35	79.5	1170 1	TSP2_BOVIN
8	35	79.5	1172 1	TSP2_HUMAN
9	35	79.5	1173 1	TSP1_XENLA
10	32	72.7	150 1	YBP2_STRAU
11	32	72.7	234 1	OPAB_NEIGO
12	32	72.7	234 1	OPAB_NEIGO
13	32	72.7	234 1	OPAF_NEIGO
14	32	72.7	260 1	OPR1_NEIMC
15	32	72.7	270 1	OPRC_NEIGO
16	32	72.7	517 1	TYR2_MOUSE
17	32	72.7	519 1	TYR2_HUMAN
18	32	72.7	522 1	TYR1_HUMAN
19	32	72.7	534 1	TYR1_ABBME
20	32	72.7	535 1	TYR1_CHICK
21	32	72.7	537 1	TYR1_HUMAN
22	32	72.7	537 1	TYR1_MOUSE
23	32	72.7	712 1	Y352_HUMAN
24	32	72.7	712 1	VGF1_IBVB
25	31	70.5	72 1	VGF1_MOUSE
26	31	70.5	72 1	VGF1_MOUSE
27	31	70.5	120 1	PA23_BOTPI
28	31	70.5	121 1	PA21_BOTPI
29	31	70.5	121 1	PA22_BOTPI
30	31	70.5	121 1	PA23_BOTPI
31	31	70.5	151 1	PA2H_BOTPI
32	31	70.5	211 1	PA2H_BOTPI
33	31	70.5	211 1	PA2H_BOTPI

34	31	70.5	211 1	TIM3_HUMAN	P35625 homo sapien
35	31	70.5	211 1	TIM3_MOUSE	P35676 mus musculu
36	31	70.5	211 1	TIM3_MOUSE	P48032 rattus norv
37	31	70.5	211 1	VE16_VACCP	P29892 vaccinia vl
38	31	70.5	236 1	PHOI_XYLFPA	09PD19 xyella fas
39	31	70.5	269 1	DAPE_CHLMT	09PD12 chlamydia m
40	31	70.5	280 1	TRDA_TREPA	083602 treponema p
41	31	70.5	306 1	YBFR_BACSU	031448 bacillus su
42	31	70.5	314 1	YDOC_SCHPO	014204 schizosacch
43	31	70.5	349 1	YCAL_YEAST	P53199 s putative
44	31	70.5	421 1	Y39E_YEAST	P47168 saccharomyc
45	31	70.5	566 1	PKNE_MYCTO	P72001 mycobacteri

## ALIGNMENTS

```

RESULT 1
TSP2_MOUSE
ID TSP2_MOUSE STANDARD: PRT; 1172 AA.
AC 003350;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT genome.";
RL J. Biol. Chem. 266:12821-12824(1991).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWF C DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: I07803; AAA53064.1; -
CC EMBL: M64866; AAA0432.1; -
CC PIR: A42587; A42587.
CC PIR: A39851; A39851.
CC HSSP: P00740; 1EDM.
CC MGD: MGI:98738; Thbs2.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR001791; Laminin_G.

```



DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR003367; tsp\_3.  
 DR Pfam; PF00008; EGF\_2.  
 DR Pfam; PF00090; tsp\_1; 3.  
 DR Pfam; PF00093; vwc; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02412; tsp\_3; 9.  
 DR SMART; SM00181; EGF; 3.  
 DR SMART; SM00209; TSP1; 3.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS00186; EGF\_2; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR PROSITE; PS01208; VWF\_C; 1.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 1 172  
 FT DOMAIN 19 172  
 FT DOMAIN 19 232  
 FT DOMAIN 318 375  
 FT DOMAIN 381 432  
 FT DOMAIN 437 493  
 FT DOMAIN 494 548  
 FT DOMAIN 549 589  
 FT DOMAIN 590 647  
 FT DOMAIN 648 692  
 FT DOMAIN 725 760  
 FT DOMAIN 761 783  
 FT DOMAIN 784 819  
 FT DOMAIN 820 842  
 FT DOMAIN 843 880  
 FT DOMAIN 881 916  
 FT DOMAIN 917 952  
 FT DOMAIN 953 1172  
 FT SITE 928 930  
 FT DISULFID 266 266  
 FT DISULFID 270 270  
 FT DISULFID 553 554  
 FT DISULFID 558 574  
 FT DISULFID 577 588  
 FT DISULFID 594 610  
 FT DISULFID 601 619  
 FT DISULFID 622 646  
 FT DISULFID 652 665  
 FT DISULFID 659 678  
 FT DISULFID 680 691  
 FT DISULFID 151 151  
 FT CARBOHYD 316 316  
 FT CARBOHYD 330 330  
 FT CARBOHYD 457 457  
 FT CARBOHYD 584 584  
 FT CARBOHYD 710 710  
 FT CARBOHYD 1069 1069  
 SQ SEQUENCE 1172 AA; 129911 MW; 7CE8E4E8599822AB CRC64;

Query Match 84.1%; Score 37; DB 1; Length 1172;  
 Best Local Similarity 75.0%; Pred. No. 10;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRWHL 8  
 DB 1108 TAYRWHLI 1115

RESULT 2  
 ID TSP2\_CHICK STANDARD; PRT: 1178 AA.  
 AC P35440;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Thrombospondin 2 precursor.  
 GN THBS2 OR TSP2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCBI\_TaxId=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91217026; PubMed-2022631;  
 RA Lawler J., Duquette M., Ferro P.;  
 RT Cloning and sequencing of chicken thrombospondin 2;  
 RL J. Biol. Chem. 266:8039-8043(1991).  
 CC - FUNCTION: ADHESIVE GLYCOPROTEIN. THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.  
 CC - SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.  
 CC - SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC - SIMILARITY: CONTRAINS 1 VWF\_C DOMAIN.  
 CC - SIMILARITY: CONTRAINS 3 EGF-LIKE DOMAINS.  
 CC - SIMILARITY: CONTRAINS 3 TSP TYPE-1 DOMAINS.  
 CC - SIMILARITY: CONTRAINS 7 TSP TYPE-3 DOMAINS.  
 CC - SIMILARITY: CONTRAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M60853; AA51437.1; -  
 DR PIR; A39804; A39804.  
 DR HSP; P00740; IEDM.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001791; Laminln\_G.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR003367; tsp\_3.  
 DR Pfam; PF00008; EGF\_1; 3.  
 DR Pfam; PF00090; tsp\_1; 3.  
 DR Pfam; PF00093; vwc; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02412; tsp\_3; 8.  
 DR Pfam; PF02412; tsp\_3; 9.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00209; TSP1; 3.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR PROSITE; PS01208; VWF\_C; 1.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 1178  
 FT DOMAIN 324 381  
 FT DOMAIN 387 438  
 FT DOMAIN 441 499  
 FT DOMAIN 500 553  
 FT DOMAIN 555 595  
 FT DOMAIN 596 653  
 FT DOMAIN 654 698  
 FT DOMAIN 731 766  
 FT DOMAIN 767 789  
 FT DOMAIN 790 825  
 FT DOMAIN 826 848  
 FT DOMAIN 849 886  
 FT POTENTIAL.  
 FT THROMBOSPONDIN 2.  
 FT HEPARIN-BINDING (POTENTIAL).  
 FT VWF\_C.  
 FT TSP TYPE-1 1.  
 FT TSP TYPE-1 2.  
 FT TSP TYPE-1 3.  
 FT EGF-LIKE 1.  
 FT EGF-LIKE 2.  
 FT EGF-LIKE 3.  
 FT CALCIUM-BINDING (POTENTIAL).  
 FT TSP TYPE-3 1.  
 FT TSP TYPE-3 2.  
 FT TSP TYPE-3 3.  
 FT TSP TYPE-3 4.  
 FT TSP TYPE-3 5.

```

FT DOMAIN 887 922 TSP TYPE-3 6.
FT DOMAIN 923 958 TSP TYPE-3 7.
FT SITE 959 1178 C-TERMINAL.
FT SITE 934 935 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 559 570 BY SIMILARITY.
FT DISULFID 564 580 BY SIMILARITY.
FT DISULFID 583 594 BY SIMILARITY.
FT DISULFID 600 616 BY SIMILARITY.
FT DISULFID 607 625 BY SIMILARITY.
FT DISULFID 628 652 BY SIMILARITY.
FT DISULFID 658 671 BY SIMILARITY.
FT DISULFID 685 697 BY SIMILARITY.
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 716 716 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1075 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1178 AA; 131816 MW; F37E02F42C8717A2 CRC64;

Query Match
Best Local Similarity 84.1%; Score 37; DB 1; Length 1178;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAYRYHL 8
DB 1114 TAYRYHL 1121

RESULT 3
CHS1_NEUCR STANDARD; PRT; 944 AA.
AC P30589;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chitin synthase 2 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
transferase 2).
GN CHS-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95039879; PubMed=7952169;
RA Dln A.B., Yarden O.;
RT "The Neurospora crassa chs-2 gene encodes a non-essential chitin
synthase."
RL Microbiology 140:2189-2197(1994).
RN [2]
RP SEQUENCE OF 250-438 FROM N.A.
RX MEDLINE=92115692; PubMed=1731323;
RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
Robbins P.W.;
RT "Classification of fungal chitin synthases."
RL Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
DE -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
-1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + ((1,4)-(N-acetyl)-
beta-D-glucosaminyl))((N) - UDP + ((1,4)-(N-acetyl)-beta-D-
glucosaminyl))((N-1)).
-1- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane.
-1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77782; CA54816.1; -
DR EMBL; M82951; AA33582.1; -
DR PIR; B45189; B45189.
DR InterPro; IPR004834; Chitin_synth.
DR Pfam; PF01644; Chitin_synth.1.
DR ProDom; PD002998; Chitin_synth.1.
KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW Multigene family.
FT TRANSMEM 597 617 POTENTIAL.
FT TRANSMEM 634 654 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 873 893 POTENTIAL.
SQ SEQUENCE 944 AA; 106816 MW; F70052AE083060D CRC64;

Query Match
Best Local Similarity 79.5%; Score 35; DB 1; Length 944;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAYRYHL 8
DB 468 SARYRYHL 475

RESULT 4
CHS1_EXODE STANDARD; PRT; 988 AA.
AC P30600; O74681.
DT 01-APR-1993 (Rel. 25, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chitin synthase 1 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
transferase 1) (Class-II chitin synthase 1).
GN CHS1.
OS Exophiala dermatitidis (Wangielia dermatitidis).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Chaetothyriomycetes;
OC Chaetothyriales; Herpotrichiellaceae; microsporid Herpotrichiellaceae;
OX NCBI_TaxID=5970;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8656;
RA Zheng L., Mendoza A.L., Wang Z., Szaniszlo P.J.;
RT "Characterization of Wdchs1, a class II chitin synthase gene, and
multiple wchs mutants of Wangielia dermatitidis."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 265-468 FROM N.A.
RX MEDLINE=92115692; PubMed=1731323;
RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
Robbins P.W.;
RT "Classification of fungal chitin synthases."
RL Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
DE -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
-1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + ((1,4)-(N-acetyl)-
beta-D-glucosaminyl))((N) - UDP + ((1,4)-(N-acetyl)-beta-D-
glucosaminyl))((N-1)).
-1- SUBCELLULAR LOCATION: Plasma membrane-bound.
-1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
II.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
DR EMBL; AF054503; AAC36064.1; -
DR EMBL; M81905; AAA30334.1; -

```

DR InterPro: IPR004834; Chitin\_synth.  
 DR Pfam: PF01644; Chitin\_synth. 1.  
 DR ProDom: PD002998; Chitin\_synth. 1.  
 KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;  
 KW Multigene family.  
 FT CONFLICT 265 CV -> KL (IN REF. 2).  
 FT CONFLICT 468 I -> L (IN REF. 2).  
 SQ SEQUENCE 988 AA; 111220 MW; 5940234B344F88A4 CRC64;

Query Match  
 Best Local Similarity 79.5%; Score 35; DB 1; Length 988;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAYRHL 8  
 DB 491 SAYRHAL 498

RESULT 5  
 CHS2\_PARB STANDARD; PRT; 1043 AA.

AC Q92444;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE Chitin synthase 2 (EC 2.4.1.16) (chitin-UDP acetyl-glucosaminyl  
 transferase 2) (Class-II chitin synthase 2).  
 GN CHS2.  
 OS Paracoccidioides brasiliensis.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Oryziales; mitosporic Onygenales; Paracoccidioides.  
 OX NCBI\_TaxID=121759;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IVIC P573 / ATCC 32071;  
 RA MEDLINE-98144800; PubMed-9483806;  
 RA Nino-Vega G.A., Buurman E.T., Goodey G.W., San-Blas G., Gow N.A.R.;  
 RT "Molecular cloning and sequencing of a chitin synthase gene (CHS2) of  
 Paracoccidioides brasiliensis.";  
 RL Yeast 14:181-187(1998).

CC -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + ((1,4)-(N-acetyl)-  
 beta-D-glucosaminyl)) (N) = UDP + ((1,4)-(N-acetyl)-beta-D-  
 glucosaminyl)) (N+1).  
 CC -1- SUBCELLULAR LOCATION: Plasma membrane-bound  
 CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS  
 II.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: Y09231; CAAT0433.1; -  
 DR InterPro: IPR004834; Chitin\_synth.  
 DR Pfam: PF01644; Chitin\_synth. 1.  
 DR ProDom: PD002998; Chitin\_synth. 1.  
 KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;  
 KW Multigene family.  
 FT TRANSMEM 663 683 POTENTIAL.  
 FT TRANSMEM 703 723 POTENTIAL.  
 FT TRANSMEM 738 758 POTENTIAL.  
 FT TRANSMEM 780 800 POTENTIAL.  
 FT TRANSMEM 907 927 POTENTIAL.  
 FT TRANSMEM 931 951 POTENTIAL.  
 SQ SEQUENCE 1043 AA; 116587 MW; F63410221B35B73 CRC64;

Query Match  
 Best Local Similarity 79.5%; Score 35; DB 1; Length 1043;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAYRHL 8  
 DB 538 SAYRHAL 545

RESULT 6  
 TSP1\_BOVIN STANDARD; PRT; 1170 AA.

AC Q28178; Q28179;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Thrombospondin 1 precursor.  
 GN TBS1 OR TSP1 OR TSP-1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Holstein; TISSUE-Tooth;  
 RX MEDLINE-98173773; PubMed-9507054;  
 RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,  
 RA Inoue H.;  
 RT "cDNA cloning of bovine thrombospondin 1 and its expression in  
 odontoblasts and predentin.";  
 RL Biochim. Biophys. Acta 1382:17-22(1998).  
 RN [2]  
 RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.  
 RC TISSUE-Aortic endothelium;  
 RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;  
 RA Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-  
 CC V/BETA-3 AND ALPHA-1IB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS  
 CC AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.  
 CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.  
 CC -1- TISSUE SPECIFICITY: ODONTOBLASTS.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AB005287; BAA21115.1; -  
 DR EMBL: X87618; CAA60950.1; -  
 DR EMBL: X87619; CAA60951.1; -  
 DR GlycoSuiteDB: Q28178; -  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF Ca.  
 DR InterPro: IPR001791; Laminin-G.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR003129; TSPN.  
 DR InterPro: IPR001037; WVF\_C.  
 DR InterPro: IPR003367; tsp\_3.  
 DR Pfam: PF00008; EGF; 2.  
 DR Pfam: PF00090; tsp\_1; 3.  
 DR Pfam: PF00093; wvc; 1.  
 DR Pfam: PF02210; TSPN; 1.  
 DR Pfam: PF02412; tsp\_3; 8.  
 DR SMART; SM00181; EGF; 3.  
 DR SMART; SM00209; TSP1; 3.

DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; WVC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01208; WVC; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 KW EGF-like domain; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 1170  
 FT DOMAIN 19 232  
 FT DOMAIN 316 373  
 FT DOMAIN 379 430  
 FT DOMAIN 435 491  
 FT DOMAIN 492 548  
 FT DOMAIN 549 587  
 FT DOMAIN 588 645  
 FT DOMAIN 646 690  
 FT DOMAIN 723 781  
 FT DOMAIN 782 817  
 FT DOMAIN 818 840  
 FT DOMAIN 841 878  
 FT DOMAIN 879 914  
 FT DOMAIN 915 950  
 FT DOMAIN 951 1170  
 FT SITE 926 928  
 FT DISULFID 270 270  
 FT DISULFID 274 274  
 FT DISULFID 351 352  
 FT DISULFID 556 572  
 FT DISULFID 575 586  
 FT DISULFID 592 608  
 FT DISULFID 599 617  
 FT DISULFID 620 644  
 FT DISULFID 650 663  
 FT DISULFID 657 676  
 FT DISULFID 678 689  
 FT CARBOHYD 248 248  
 FT CARBOHYD 360 360  
 FT CARBOHYD 708 708  
 FT CARBOHYD 1067 1067  
 FT CARBOHYD 1085 1085  
 FT CONFLICT 805 805  
 SO SEQUENCE 1170 AA; 129533 MW; 0DD6AD355FA031A CMC64;  
 Query Match 79.5%; Score 35; DB 1; Length 1170;  
 Best Local Similarity 85.7%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAYRHL 7  
 DB 1106 TAYRHL 1112

RESULT 7  
 TSP2\_BOVIN STANDARD; PRT; 1170 AA.  
 AC 095116; Q28180;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Thrombospondin 2 precursor (Corticotropic-induced secreted protein)  
 DE (CISF).  
 GN THBS2 OR TSP2 OR TSP-2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dank M., Chinn A., Lafeuillade M., Karamidas M., Aguesse-Gernon S.,

RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-522 FROM N.A.  
 RX MEDLINE=96331130; PUBMED=8698834;  
 RA Lafeuillade B., Pelletier S., Karamidas M., Dank M., Chambaz E.M.,  
 RA Feige J.J.;  
 RT "Opposite regulation of thrombospondin-1 and corticotropin-induced  
 RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic  
 RT hormone in adrenocortical cells."  
 RT J. Cell. Physiol. 167:164-172(1996).  
 RL [3]  
 RN SEQUENCE OF 318-831 FROM N.A.  
 RP TISSUE-Aortic endothelium;  
 RC Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;  
 RA "Cloning and sequencing of bovine thrombospondin stimulatory effect of  
 RT TGF-beta."  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC -1- SUBUNIT: HOMOTRIMER: DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 WVC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X96540; CAA65385.1; -  
 DR EMBL; X87620; CAA60952.1; -  
 DR HSSP; P00740; 1EDM.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR003129; TSPN.  
 DR InterPro: IPR001007; WVC\_C.  
 DR InterPro: IPR003367; TSP\_3.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00090; TSP\_1; 3.  
 DR Pfam; PF00093; WVC; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02412; TSP\_3; 9.  
 DR SMART; SM00181; EGF; 3.  
 DR SMART; SM00209; TSP1; 3.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; WVC; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR PROSITE; PS01208; WVC; 1.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 KW EGF-like domain; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 1170  
 FT DOMAIN 19 232  
 FT DOMAIN 316 373  
 FT DOMAIN 379 430  
 FT DOMAIN 435 491  
 FT DOMAIN 492 546  
 FT DOMAIN 547 587  
 FT DOMAIN 588 645  
 FT DOMAIN 646 690  
 FT DOMAIN 723 758  
 FT DOMAIN 759 781  
 FT POTENTIAL.  
 FT THROMBOSPONDIN 2.  
 FT HEPARIN-BINDING (POTENTIAL).  
 FT WVC.  
 FT TSP TYPE-1 1.  
 FT TSP TYPE-1 2.  
 FT TSP TYPE-1 3.  
 FT EGF-LIKE 1.  
 FT EGF-LIKE 2.  
 FT EGF-LIKE 3.  
 FT EGF-LIKE 4.  
 FT EGF-LIKE 5.  
 FT EGF-LIKE 6.  
 FT EGF-LIKE 7.  
 FT EGF-LIKE 8.  
 FT EGF-LIKE 9.  
 FT EGF-LIKE 10.  
 FT EGF-LIKE 11.  
 FT EGF-LIKE 12.  
 FT EGF-LIKE 13.  
 FT EGF-LIKE 14.  
 FT EGF-LIKE 15.  
 FT EGF-LIKE 16.  
 FT EGF-LIKE 17.  
 FT EGF-LIKE 18.  
 FT EGF-LIKE 19.  
 FT EGF-LIKE 20.  
 FT EGF-LIKE 21.  
 FT EGF-LIKE 22.  
 FT EGF-LIKE 23.  
 FT EGF-LIKE 24.  
 FT EGF-LIKE 25.  
 FT EGF-LIKE 26.  
 FT EGF-LIKE 27.  
 FT EGF-LIKE 28.  
 FT EGF-LIKE 29.  
 FT EGF-LIKE 30.  
 FT EGF-LIKE 31.  
 FT EGF-LIKE 32.  
 FT EGF-LIKE 33.  
 FT EGF-LIKE 34.  
 FT EGF-LIKE 35.  
 FT EGF-LIKE 36.  
 FT EGF-LIKE 37.  
 FT EGF-LIKE 38.  
 FT EGF-LIKE 39.  
 FT EGF-LIKE 40.  
 FT EGF-LIKE 41.  
 FT EGF-LIKE 42.  
 FT EGF-LIKE 43.  
 FT EGF-LIKE 44.  
 FT EGF-LIKE 45.  
 FT EGF-LIKE 46.  
 FT EGF-LIKE 47.  
 FT EGF-LIKE 48.  
 FT EGF-LIKE 49.  
 FT EGF-LIKE 50.  
 FT EGF-LIKE 51.  
 FT EGF-LIKE 52.  
 FT EGF-LIKE 53.  
 FT EGF-LIKE 54.  
 FT EGF-LIKE 55.  
 FT EGF-LIKE 56.  
 FT EGF-LIKE 57.  
 FT EGF-LIKE 58.  
 FT EGF-LIKE 59.  
 FT EGF-LIKE 60.  
 FT EGF-LIKE 61.  
 FT EGF-LIKE 62.  
 FT EGF-LIKE 63.  
 FT EGF-LIKE 64.  
 FT EGF-LIKE 65.  
 FT EGF-LIKE 66.  
 FT EGF-LIKE 67.  
 FT EGF-LIKE 68.  
 FT EGF-LIKE 69.  
 FT EGF-LIKE 70.  
 FT EGF-LIKE 71.  
 FT EGF-LIKE 72.  
 FT EGF-LIKE 73.  
 FT EGF-LIKE 74.  
 FT EGF-LIKE 75.  
 FT EGF-LIKE 76.  
 FT EGF-LIKE 77.  
 FT EGF-LIKE 78.  
 FT EGF-LIKE 79.  
 FT EGF-LIKE 80.  
 FT EGF-LIKE 81.  
 FT EGF-LIKE 82.  
 FT EGF-LIKE 83.  
 FT EGF-LIKE 84.  
 FT EGF-LIKE 85.  
 FT EGF-LIKE 86.  
 FT EGF-LIKE 87.  
 FT EGF-LIKE 88.  
 FT EGF-LIKE 89.  
 FT EGF-LIKE 90.  
 FT EGF-LIKE 91.  
 FT EGF-LIKE 92.  
 FT EGF-LIKE 93.  
 FT EGF-LIKE 94.  
 FT EGF-LIKE 95.  
 FT EGF-LIKE 96.  
 FT EGF-LIKE 97.  
 FT EGF-LIKE 98.  
 FT EGF-LIKE 99.  
 FT EGF-LIKE 100.  
 FT EGF-LIKE 101.  
 FT EGF-LIKE 102.  
 FT EGF-LIKE 103.  
 FT EGF-LIKE 104.  
 FT EGF-LIKE 105.  
 FT EGF-LIKE 106.  
 FT EGF-LIKE 107.  
 FT EGF-LIKE 108.  
 FT EGF-LIKE 109.  
 FT EGF-LIKE 110.  
 FT EGF-LIKE 111.  
 FT EGF-LIKE 112.  
 FT EGF-LIKE 113.  
 FT EGF-LIKE 114.  
 FT EGF-LIKE 115.  
 FT EGF-LIKE 116.  
 FT EGF-LIKE 117.  
 FT EGF-LIKE 118.  
 FT EGF-LIKE 119.  
 FT EGF-LIKE 120.  
 FT EGF-LIKE 121.  
 FT EGF-LIKE 122.  
 FT EGF-LIKE 123.  
 FT EGF-LIKE 124.  
 FT EGF-LIKE 125.  
 FT EGF-LIKE 126.  
 FT EGF-LIKE 127.  
 FT EGF-LIKE 128.  
 FT EGF-LIKE 129.  
 FT EGF-LIKE 130.  
 FT EGF-LIKE 131.  
 FT EGF-LIKE 132.  
 FT EGF-LIKE 133.  
 FT EGF-LIKE 134.  
 FT EGF-LIKE 135.  
 FT EGF-LIKE 136.  
 FT EGF-LIKE 137.  
 FT EGF-LIKE 138.  
 FT EGF-LIKE 139.  
 FT EGF-LIKE 140.  
 FT EGF-LIKE 141.  
 FT EGF-LIKE 142.  
 FT EGF-LIKE 143.  
 FT EGF-LIKE 144.  
 FT EGF-LIKE 145.  
 FT EGF-LIKE 146.  
 FT EGF-LIKE 147.  
 FT EGF-LIKE 148.  
 FT EGF-LIKE 149.  
 FT EGF-LIKE 150.  
 FT EGF-LIKE 151.  
 FT EGF-LIKE 152.  
 FT EGF-LIKE 153.  
 FT EGF-LIKE 154.  
 FT EGF-LIKE 155.  
 FT EGF-LIKE 156.  
 FT EGF-LIKE 157.  
 FT EGF-LIKE 158.  
 FT EGF-LIKE 159.  
 FT EGF-LIKE 160.  
 FT EGF-LIKE 161.  
 FT EGF-LIKE 162.  
 FT EGF-LIKE 163.  
 FT EGF-LIKE 164.  
 FT EGF-LIKE 165.  
 FT EGF-LIKE 166.  
 FT EGF-LIKE 167.  
 FT EGF-LIKE 168.  
 FT EGF-LIKE 169.  
 FT EGF-LIKE 170.  
 FT EGF-LIKE 171.  
 FT EGF-LIKE 172.  
 FT EGF-LIKE 173.  
 FT EGF-LIKE 174.  
 FT EGF-LIKE 175.  
 FT EGF-LIKE 176.  
 FT EGF-LIKE 177.  
 FT EGF-LIKE 178.  
 FT EGF-LIKE 179.  
 FT EGF-LIKE 180.  
 FT EGF-LIKE 181.  
 FT EGF-LIKE 182.  
 FT EGF-LIKE 183.  
 FT EGF-LIKE 184.  
 FT EGF-LIKE 185.  
 FT EGF-LIKE 186.  
 FT EGF-LIKE 187.  
 FT EGF-LIKE 188.  
 FT EGF-LIKE 189.  
 FT EGF-LIKE 190.  
 FT EGF-LIKE 191.  
 FT EGF-LIKE 192.  
 FT EGF-LIKE 193.  
 FT EGF-LIKE 194.  
 FT EGF-LIKE 195.  
 FT EGF-LIKE 196.  
 FT EGF-LIKE 197.  
 FT EGF-LIKE 198.  
 FT EGF-LIKE 199.  
 FT EGF-LIKE 200.  
 FT EGF-LIKE 201.  
 FT EGF-LIKE 202.  
 FT EGF-LIKE 203.  
 FT EGF-LIKE 204.  
 FT EGF-LIKE 205.  
 FT EGF-LIKE 206.  
 FT EGF-LIKE 207.  
 FT EGF-LIKE 208.  
 FT EGF-LIKE 209.  
 FT EGF-LIKE 210.  
 FT EGF-LIKE 211.  
 FT EGF-LIKE 212.  
 FT EGF-LIKE 213.  
 FT EGF-LIKE 214.  
 FT EGF-LIKE 215.  
 FT EGF-LIKE 216.  
 FT EGF-LIKE 217.  
 FT EGF-LIKE 218.  
 FT EGF-LIKE 219.  
 FT EGF-LIKE 220.  
 FT EGF-LIKE 221.  
 FT EGF-LIKE 222.  
 FT EGF-LIKE 223.  
 FT EGF-LIKE 224.  
 FT EGF-LIKE 225.  
 FT EGF-LIKE 226.  
 FT EGF-LIKE 227.  
 FT EGF-LIKE 228.  
 FT EGF-LIKE 229.  
 FT EGF-LIKE 230.  
 FT EGF-LIKE 231.  
 FT EGF-LIKE 232.  
 FT EGF-LIKE 233.  
 FT EGF-LIKE 234.  
 FT EGF-LIKE 235.  
 FT EGF-LIKE 236.  
 FT EGF-LIKE 237.  
 FT EGF-LIKE 238.  
 FT EGF-LIKE 239.  
 FT EGF-LIKE 240.  
 FT EGF-LIKE 241.  
 FT EGF-LIKE 242.  
 FT EGF-LIKE 243.  
 FT EGF-LIKE 244.  
 FT EGF-LIKE 245.  
 FT EGF-LIKE 246.  
 FT EGF-LIKE 247.  
 FT EGF-LIKE 248.  
 FT EGF-LIKE 249.  
 FT EGF-LIKE 250.  
 FT EGF-LIKE 251.  
 FT EGF-LIKE 252.  
 FT EGF-LIKE 253.  
 FT EGF-LIKE 254.  
 FT EGF-LIKE 255.  
 FT EGF-LIKE 256.  
 FT EGF-LIKE 257.  
 FT EGF-LIKE 258.  
 FT EGF-LIKE 259.  
 FT EGF-LIKE 260.  
 FT EGF-LIKE 261.  
 FT EGF-LIKE 262.  
 FT EGF-LIKE 263.  
 FT EGF-LIKE 264.  
 FT EGF-LIKE 265.  
 FT EGF-LIKE 266.  
 FT EGF-LIKE 267.  
 FT EGF-LIKE 268.  
 FT EGF-LIKE 269.  
 FT EGF-LIKE 270.  
 FT EGF-LIKE 271.  
 FT EGF-LIKE 272.  
 FT EGF-LIKE 273.  
 FT EGF-LIKE 274.  
 FT EGF-LIKE 275.  
 FT EGF-LIKE 276.  
 FT EGF-LIKE 277.  
 FT EGF-LIKE 278.  
 FT EGF-LIKE 279.  
 FT EGF-LIKE 280.  
 FT EGF-LIKE 281.  
 FT EGF-LIKE 282.  
 FT EGF-LIKE 283.  
 FT EGF-LIKE 284.  
 FT EGF-LIKE 285.  
 FT EGF-LIKE 286.  
 FT EGF-LIKE 287.  
 FT EGF-LIKE 288.  
 FT EGF-LIKE 289.  
 FT EGF-LIKE 290.  
 FT EGF-LIKE 291.  
 FT EGF-LIKE 292.  
 FT EGF-LIKE 293.  
 FT EGF-LIKE 294.  
 FT EGF-LIKE 295.  
 FT EGF-LIKE 296.  
 FT EGF-LIKE 297.  
 FT EGF-LIKE 298.  
 FT EGF-LIKE 299.  
 FT EGF-LIKE 300.  
 FT EGF-LIKE 301.  
 FT EGF-LIKE 302.  
 FT EGF-LIKE 303.  
 FT EGF-LIKE 304.  
 FT EGF-LIKE 305.  
 FT EGF-LIKE 306.  
 FT EGF-LIKE 307.  
 FT EGF-LIKE 308.  
 FT EGF-LIKE 309.  
 FT EGF-LIKE 310.  
 FT EGF-LIKE 311.  
 FT EGF-LIKE 312.  
 FT EGF-LIKE 313.  
 FT EGF-LIKE 314.  
 FT EGF-LIKE 315.  
 FT EGF-LIKE 316.  
 FT EGF-LIKE 317.  
 FT EGF-LIKE 318.  
 FT EGF-LIKE 319.  
 FT EGF-LIKE 320.  
 FT EGF-LIKE 321.  
 FT EGF-LIKE 322.  
 FT EGF-LIKE 323.  
 FT EGF-LIKE 324.  
 FT EGF-LIKE 325.  
 FT EGF-LIKE 326.  
 FT EGF-LIKE 327.  
 FT EGF-LIKE 328.  
 FT EGF-LIKE 329.  
 FT EGF-LIKE 330.  
 FT EGF-LIKE 331.  
 FT EGF-LIKE 332.  
 FT EGF-LIKE 333.  
 FT EGF-LIKE 334.  
 FT EGF-LIKE 335.  
 FT EGF-LIKE 336.  
 FT EGF-LIKE 337.  
 FT EGF-LIKE 338.  
 FT EGF-LIKE 339.  
 FT EGF-LIKE 340.  
 FT EGF-LIKE 341.  
 FT EGF-LIKE 342.  
 FT EGF-LIKE 343.  
 FT EGF-LIKE 344.  
 FT EGF-LIKE 345.  
 FT EGF-LIKE 346.  
 FT EGF-LIKE 347.  
 FT EGF-LIKE 348.  
 FT EGF-LIKE 349.  
 FT EGF-LIKE 350.  
 FT EGF-LIKE 351.  
 FT EGF-LIKE 352.  
 FT EGF-LIKE 353.  
 FT EGF-LIKE 354.  
 FT EGF-LIKE 355.  
 FT EGF-LIKE 356.  
 FT EGF-LIKE 357.  
 FT EGF-LIKE 358.  
 FT EGF-LIKE 359.  
 FT EGF-LIKE 360.  
 FT EGF-LIKE 361.  
 FT EGF-LIKE 362.  
 FT EGF-LIKE 363.  
 FT EGF-LIKE 364.  
 FT EGF-LIKE 365.  
 FT EGF-LIKE 366.  
 FT EGF-LIKE 367.  
 FT EGF-LIKE 368.  
 FT EGF-LIKE 369.  
 FT EGF-LIKE 370.  
 FT EGF-LIKE 371.  
 FT EGF-LIKE 372.  
 FT EGF-LIKE 373.  
 FT EGF-LIKE 374.  
 FT EGF-LIKE 375.  
 FT EGF-LIKE 376.  
 FT EGF-LIKE 377.  
 FT EGF-LIKE 378.  
 FT EGF-LIKE 379.  
 FT EGF-LIKE 380.  
 FT EGF-LIKE 381.  
 FT EGF-LIKE 382.  
 FT EGF-LIKE 383.  
 FT EGF-LIKE 384.  
 FT EGF-LIKE 385.  
 FT EGF-LIKE 386.  
 FT EGF-LIKE 387.  
 FT EGF-LIKE 388.  
 FT EGF-LIKE 389.  
 FT EGF-LIKE 390.  
 FT EGF-LIKE 391.  
 FT EGF-LIKE 392.  
 FT EGF-LIKE 393.  
 FT EGF-LIKE 394.  
 FT EGF-LIKE 395.  
 FT EGF-LIKE 396.  
 FT EGF-LIKE 397.  
 FT EGF-LIKE 398.  
 FT EGF-LIKE 399.  
 FT EGF-LIKE 400.  
 FT EGF-LIKE 401.  
 FT EGF-LIKE 402.  
 FT EGF-LIKE 403.  
 FT EGF-LIKE 404.  
 FT EGF-LIKE 405.  
 FT EGF-LIKE 406.  
 FT EGF-LIKE 407.  
 FT EGF-LIKE 408.  
 FT EGF-LIKE 409.  
 FT EGF-LIKE 410.  
 FT EGF-LIKE 411.  
 FT EGF-LIKE 412.  
 FT EGF-LIKE 413.  
 FT EGF-LIKE 414.  
 FT EGF-LIKE 415.  
 FT EGF-LIKE 416.  
 FT EGF-LIKE 417.  
 FT EGF-LIKE 418.  
 FT EGF-LIKE 419.  
 FT EGF-LIKE 420.  
 FT EGF-LIKE 421.  
 FT EGF-LIKE 422.  
 FT EGF-LIKE 423.  
 FT EGF-LIKE 424.  
 FT EGF-LIKE 425.  
 FT EGF-LIKE 426.  
 FT EGF-LIKE 427.  
 FT EGF-LIKE 428.  
 FT EGF-LIKE 429.  
 FT EGF-LIKE 430.  
 FT EGF-LIKE 431.  
 FT EGF-LIKE 432.  
 FT EGF-LIKE 433.  
 FT EGF-LIKE 434.  
 FT EGF-LIKE 435.  
 FT EGF-LIKE 436.  
 FT EGF-LIKE 437.  
 FT EGF-LIKE 438.  
 FT EGF-LIKE 439.  
 FT EGF-LIKE 440.  
 FT EGF-LIKE 441.  
 FT EGF-LIKE 442.  
 FT EGF-LIKE 443.  
 FT EGF-LIKE 444.  
 FT EGF-LIKE 445.  
 FT EGF-LIKE 446.  
 FT EGF-LIKE 447.  
 FT EGF-LIKE 448.  
 FT EGF-LIKE 449.  
 FT EGF-LIKE 450.  
 FT EGF-LIKE 451.  
 FT EGF-LIKE 452.  
 FT EGF-LIKE 453.  
 FT EGF-LIKE 454.  
 FT EGF-LIKE 455.  
 FT EGF-LIKE 456.  
 FT EGF-LIKE 457.  
 FT EGF-LIKE 458.  
 FT EGF-LIKE 459.  
 FT EGF-LIKE 460.  
 FT EGF-LIKE 461.  
 FT EGF-LIKE 462.  
 FT EGF-LIKE 463.  
 FT EGF-LIKE 464.  
 FT EGF-LIKE 465.  
 FT EGF-LIKE 466.  
 FT EGF-LIKE 467.  
 FT EGF-LIKE 468.  
 FT EGF-LIKE 469.  
 FT EGF-LIKE 470.  
 FT EGF-LIKE 471.  
 FT EGF-LIKE 472.  
 FT EGF-LIKE 473.  
 FT EGF-LIKE 474.  
 FT EGF-LIKE 475.  
 FT EGF-LIKE 476.  
 FT EGF-LIKE 477.  
 FT EGF-LIKE 478.  
 FT EGF-LIKE 479.  
 FT EGF-LIKE 480.  
 FT EGF-LIKE 481.  
 FT EGF-LIKE 482.  
 FT EGF-LIKE 483.  
 FT EGF-LIKE 484.  
 FT EGF-LIKE 485.  
 FT EGF-LIKE 486.  
 FT EGF-LIKE 487.  
 FT EGF-LIKE 488.  
 FT EGF-LIKE 489.  
 FT EGF-LIKE 490.  
 FT EGF-LIKE 491.  
 FT EGF-LIKE 492.  
 FT EGF-LIKE 493.  
 FT EGF-LIKE 494.  
 FT EGF-LIKE 495.  
 FT EGF-LIKE 496.  
 FT EGF-LIKE 497.  
 FT EGF-LIKE 498.  
 FT EGF-LIKE 499.  
 FT EGF-LIKE 500.  
 FT EGF-LIKE 501.  
 FT EGF-LIKE 502.  
 FT EGF-LIKE 503.  
 FT EGF-LIKE 504.  
 FT EGF-LIKE 505.  
 FT EGF-LIKE 506.  
 FT EGF-LIKE 507.  
 FT EGF-LIKE 508.  
 FT EGF-LIKE 509.  
 FT EGF-LIKE 510.  
 FT EGF-LIKE 511.  
 FT EGF-LIKE 512.  
 FT EGF-LIKE 513.  
 FT EGF-LIKE 514.  
 FT EGF-LIKE 515.  
 FT EGF-LIKE 516.  
 FT EGF-LIKE 517.  
 FT EGF-LIKE 518.  
 FT EGF-LIKE 519.  
 FT EGF-LIKE 520.  
 FT EGF-LIKE 521.  
 FT EGF-LIKE 522.  
 FT EGF-LIKE 523.  
 FT EGF-LIKE 524.  
 FT EGF-LIKE 525.  
 FT EGF-LIKE 526.  
 FT EGF-LIKE 527.  
 FT EGF-LIKE 528.  
 FT EGF-LIKE 529.  
 FT EGF-LIKE 530.  
 FT EGF-LIKE 531.  
 FT EGF-LIKE 532.  
 FT EGF-LIKE 533.  
 FT EGF-LIKE 534.  
 FT EGF-LIKE 535.  
 FT EGF-LIKE 536.  
 FT EGF-LIKE 537.  
 FT EGF-LIKE 538.  
 FT EGF-LIKE 539.  
 FT EGF-LIKE 540.  
 FT EGF-LIKE 541.  
 FT EGF-LIKE 542.  
 FT EGF-LIKE 543.  
 FT EGF-LIKE 544.  
 FT EGF-LIKE 545.  
 FT EGF-LIKE 546.  
 FT EGF-LIKE 547.  
 FT EGF-LIKE 548.  
 FT EGF-LIKE 549.  
 FT EGF-LIKE 550.  
 FT EGF-LIKE 551.  
 FT EGF-LIKE 552.  
 FT EGF-LIKE 553.  
 FT EGF-LIKE 554.  
 FT EGF-LIKE 555.  
 FT EGF-LIKE 556.  
 FT EGF-LIKE 557.  
 FT EGF-LIKE 558.  
 FT EGF-LIKE 559.  
 FT EGF-LIKE 560.  
 FT EGF-LIKE 561.  
 FT EGF-LIKE 562.  
 FT EGF-LIKE 563.  
 FT EGF-LIKE 564.  
 FT EGF-LIKE 565.  
 FT EGF-LIKE 566.  
 FT EGF-LIKE 567.  
 FT EGF-LIKE 568.  
 FT EGF-LIKE 569.  
 FT EGF-LIKE 570.  
 FT EGF-LIKE 571.  
 FT EGF-LIKE 572.  
 FT EGF-LIKE 573.  
 FT EGF-LIKE 574.  
 FT EGF-LIKE 575.  
 FT EGF-LIKE 576.  
 FT EGF-LIKE 577.  
 FT EGF-LIKE 578.  
 FT EGF-LIKE 579.  
 FT EGF-LIKE 580.  
 FT EGF-LIKE 581.  
 FT EGF-LIKE 582.  
 FT EGF-LIKE 583.  
 FT EGF-LIKE 584.  
 FT EGF-LIKE 585.  
 FT EGF-LIKE 586.  
 FT EGF-LIKE 587.  
 FT EGF-LIKE 588.  
 FT EGF-LIKE 589.  
 FT EGF-LIKE 590.  
 FT EGF-LIKE 591.  
 FT EGF-LIKE 592.  
 FT EGF-LIKE 593.  
 FT EGF-LIKE 594.  
 FT EGF-LIKE 595.  
 FT EGF-LIKE 596.  
 FT EGF-LIKE 597.  
 FT EGF-LIKE 598.  
 FT EGF-LIKE 599.  
 FT EGF-LIKE 600.  
 FT EGF-LIKE 601.  
 FT EGF-LIKE 602.  
 FT EGF-LIKE 603.  
 FT EGF-LIKE 604.  
 FT EGF-LIKE 605.  
 FT EGF-LIKE 606.  
 FT EGF-LIKE 607.  
 FT EGF-LIKE 608.  
 FT EGF-LIKE 609.  
 FT EGF-LIKE 610.  
 FT EGF-LIKE 611.  
 FT EGF-LIKE 612.  
 FT EGF-LIKE 613.  
 FT EGF-LIKE 614.  
 FT EGF-LIKE 615.  
 FT EGF-LIKE 616.  
 FT EGF-LIKE 617.  
 FT EGF-LIKE 618.  
 FT EGF-LIKE 619.  
 FT EGF-LIKE 620.  
 FT EGF-LIKE 621.  
 FT EGF-LIKE 622.  
 FT EGF-LIKE 623.  
 FT EGF-LIKE 624.  
 FT EGF-LIKE 625.  
 FT EGF-LIKE 626.  
 FT EGF-LIKE 627.  
 FT EGF-LIKE 628.  
 FT EGF-LIKE 629.  
 FT EGF-LIKE 630.  
 FT EGF-LIKE 631.  
 FT EGF-LIKE 632.  
 FT EGF-LIKE 633.  
 FT EGF-LIKE 634.  
 FT EGF-LIKE 635.  
 FT EGF-LIKE 636.  
 FT EGF-LIKE 637.  
 FT EGF-LIKE 638.  
 FT EGF-LIKE 639.  
 FT EGF-LIKE 640.  
 FT EGF-LIKE 641.  
 FT EGF-LIKE 642.  
 FT EGF-LIKE 643.  
 FT EGF-LIKE 644.  
 FT EGF-LIKE 645.  
 FT EGF-LIKE 646.  
 FT EGF-LIKE 647.  
 FT EGF-LIKE 648.  
 FT EGF-LIKE 649.  
 FT EGF-LIKE 650.  
 FT EGF-LIKE 651.  
 FT EGF-LIKE 652.  
 FT EGF-LIKE 653.  
 FT EGF-LIKE 654.  
 FT EGF-LIKE 655.  
 FT EGF-LIKE 656.  
 FT EGF-LIKE 657.  
 FT EGF-LIKE 658.  
 FT EGF-LIKE 659.  
 FT EGF-LIKE 660.  
 FT EGF-LIKE 661.  
 FT EGF-LIKE 662.  
 FT EGF-LIKE 663.  
 FT EGF-LIKE 664.  
 FT EGF-LIKE 665.  
 FT EGF-LIKE 666.  
 FT EGF-LIKE 667.  
 FT EGF-LIKE 668.  
 FT EGF-LIKE 669.  
 FT EGF-LIKE 670.  
 FT EGF-LIKE 671.  
 FT EGF-LIKE 672.  
 FT EGF-LIKE 673.  
 FT EGF-LIKE 674.  
 FT EGF-LIKE 675.  
 FT EGF-LIKE 676.  
 FT EGF-LIKE 677.  
 FT EGF-LIKE 678.  
 FT EGF-LIKE 679.  
 FT EGF-LIKE 680.  
 FT EGF-LIKE 681.  
 FT EGF-LIKE 682.

FT DOMAIN 782 817 TSP TYPE-3 3.  
 FT DOMAIN 818 840 TSP TYPE-3 4.  
 FT DOMAIN 841 878 TSP TYPE-3 5.  
 FT DOMAIN 879 914 TSP TYPE-3 6.  
 FT DOMAIN 915 950 TSP TYPE-3 7.  
 FT DOMAIN 951 1170 C-TERMINAL.  
 FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).  
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).  
 FT DISULFID 551 562 BY SIMILARITY.  
 FT DISULFID 556 572 BY SIMILARITY.  
 FT DISULFID 575 586 BY SIMILARITY.  
 FT DISULFID 592 608 BY SIMILARITY.  
 FT DISULFID 599 617 BY SIMILARITY.  
 FT DISULFID 620 644 BY SIMILARITY.  
 FT DISULFID 650 663 BY SIMILARITY.  
 FT DISULFID 657 676 BY SIMILARITY.  
 FT CARBOHYD 678 689 BY SIMILARITY.  
 FT CARBOHYD 151 151 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 316 316 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 435 455 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 582 582 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 708 708 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 936 936 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CARBOHYD 1067 1067 N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 FT CONFLICT 535 535 A -> V (IN REF. 3).  
 FT CONFLICT 748 748 S -> T (IN REF. 3).  
 SQ SEQUENCE 1170 AA, 129862 MW, 9CFL1BF5589A051 CRC64;

Query Match 79.5%; Score 35; DB 1; Length 1170;  
 Best Local Similarity 85.7%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRHL 7  
 DB 1106 TAYRHL 1112

RESULT 8  
 TSP2\_HUMAN STANDARD; PRT: 1172 AA.  
 ID TSP2\_HUMAN  
 AC P35442; 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN Thrombospondin 2 precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94010892; PubMed=8406456;  
 RA Label T.L., Byers P.H.;  
 RT "Sequence and characterization of the complete human thrombospondin 2  
 cDNA: potential regulatory role for the 3' untranslated region.";  
 RL Genomics 17:225-229(1993).  
 RN [2]  
 RP SEQUENCE OF 560-1172 FROM N.A.  
 RX TISSUE=Plasmodium;  
 RC MEDLINE=92217961; PubMed=1559694;  
 RA Label T.L., McGookey Milwicz D.J., Distche C.M., Byers P.H.;  
 RT "Thrombospondin II: partial cDNA sequence, chromosome location, and  
 expression of a second member of the thrombospondin gene family in  
 humans";  
 RL Genomics 12:421-429(1992).  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC -1- SUBUNIT: HOMOTRIMER. DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.sdb.ch/announce/>  
 CC or send an email to [license@isb.sdb.ch](mailto:license@isb.sdb.ch)).  
 CC -----  
 CC EMBL, L12350; AAA03703.1;  
 CC EMBL, M81339; NOT\_ANNOTATED\_CDS.  
 CC PIR, A42173; A42173.  
 CC HSP, P00740; 1EDM.  
 CC Genew; HGNC:11786; THBS2.  
 CC MIM; 180061;  
 CC InterPro; IPR000561; EGF-like.  
 CC InterPro; IPR001881; EGF-Ca.  
 CC InterPro; IPR001791; laminin-G.  
 CC InterPro; IPR000884; TSP1.  
 CC InterPro; IPR003129; TSPN.  
 CC InterPro; IPR001007; VWFC.  
 CC InterPro; IPR003367; tsp\_3.  
 CC Pfam; PF00008; EGF; 2.  
 CC Pfam; PF00090; tsp\_1; 3.  
 CC Pfam; PF00093; vwc; 1.  
 CC Pfam; PF02210; TSPN; 1.  
 CC Pfam; PF02412; tsp\_3; 9.  
 CC SMART; SM00181; EGF; 3.  
 CC SMART; SM00209; TSP1; 3.  
 CC SMART; SM00210; TSPN; 1.  
 CC SMART; SM00214; vwc; 1.  
 CC PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 CC PROSITE; PS01186; EGF\_2; 1.  
 CC PROSITE; PS00092; TSP1; 3.  
 CC PROSITE; PS01208; VWFC; 1.  
 CC KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 CC EGF-like domain; Signal.  
 CC FT SIGNAL 1 18  
 CC FT CHAIN 19 1172  
 CC FT DOMAIN 19 232  
 CC FT DOMAIN 318 375  
 CC FT DOMAIN 381 432  
 CC FT DOMAIN 437 493  
 CC FT DOMAIN 494 548  
 CC FT DOMAIN 549 589  
 CC FT DOMAIN 590 647  
 CC FT DOMAIN 647 692  
 CC FT DOMAIN 692 725  
 CC FT DOMAIN 725 760  
 CC FT DOMAIN 761 783  
 CC FT DOMAIN 784 819  
 CC FT DOMAIN 820 842  
 CC FT DOMAIN 843 880  
 CC FT DOMAIN 881 916  
 CC FT DOMAIN 916 952  
 CC FT DOMAIN 952 972  
 CC FT SITE 926 930  
 CC FT DISULFID 266 266  
 CC FT DISULFID 270 270  
 CC FT DISULFID 553 564  
 CC FT DISULFID 558 574  
 CC FT DISULFID 577 588  
 CC FT DISULFID 594 610  
 CC FT DISULFID 601 619  
 CC FT DISULFID 622 646  
 CC FT DISULFID 652 665  
 CC FT DISULFID 659 678  
 CC FT DISULFID 680 691  
 CC FT CARBOHYD 151 151  
 CC FT CARBOHYD 316 316  
 CC -----  
 CC POTENTIAL.  
 CC THROMBOSPONDIN 2.  
 CC HEPARIN-BINDING (POTENTIAL).  
 CC VWFC.  
 CC TSP TYPE-1 1.  
 CC TSP TYPE-1 2.  
 CC TSP TYPE-1 3.  
 CC EGF-LIKE 1.  
 CC EGF-LIKE 2.  
 CC EGF-LIKE 3.  
 CC TSP TYPE-3 1.  
 CC TSP TYPE-3 2.  
 CC TSP TYPE-3 3.  
 CC TSP TYPE-3 4.  
 CC TSP TYPE-3 5.  
 CC TSP TYPE-3 6.  
 CC TSP TYPE-3 7.  
 CC C-TERMINAL.  
 CC CELL ATTACHMENT SITE (POTENTIAL).  
 CC INTERCHAIN (PROBABLE).  
 CC INTERCHAIN (PROBABLE).  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC N-LINKED (GLCNCNC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNCNC. . .) (POTENTIAL).

```

FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB30E44C6F5 CR664;

Query Match 79.5%; Score 35; DB 1; Length 1172;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRHL 7
Db 1108 TAYRWHL 1114

RESULT 9
TSPL_XENLA STANDARD; PRT; 1173 AA.
ID TSPL_XENLA
AC P35448;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Urry L.A., Ramos J., Duquette M., Desimone D.W., Lawler J.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
V/BETA-3 AND ALPHA-1IB/BETA-3 (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWF-C DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04278; -. NOT_ANNOTATED_CDS.
CC HSSP: P00740; 1EDM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR000884; TSPL.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; VWF_C.
DR InterPro: IPR003367; tsp_3.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00093; tsp_1; 3.
DR Pfam: PF02210; TSPN_1.
DR Pfam: PF02412; tsp_3; 8.
DR SMART: SM00181; EGF_2.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00209; TSP1_3.
DR SMART: SM00210; TSPN_1.
DR SMART: SM00214; VMC_1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.

```

```

DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS50092; TSP1_3.
DR PROSITE: PS01208; VWF_C; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
EGF-like domain; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1173
FT DOMAIN 23 235
FT DOMAIN 319 376
FT DOMAIN 382 433
FT DOMAIN 438 494
FT DOMAIN 495 546
FT DOMAIN 550 590
FT DOMAIN 591 648
FT DOMAIN 649 693
FT DOMAIN 726 761
FT DOMAIN 762 784
FT DOMAIN 785 820
FT DOMAIN 821 843
FT DOMAIN 844 881
FT DOMAIN 882 917
FT DOMAIN 918 953
FT DOMAIN 954 1173
FT SITE 929 931
FT DISULFID 554 565
FT DISULFID 559 575
FT DISULFID 578 589
FT DISULFID 595 611
FT DISULFID 602 620
FT DISULFID 623 647
FT DISULFID 653 666
FT DISULFID 660 679
FT DISULFID 681 692
FT CARBOHYD 155 155
FT CARBOHYD 158 158
FT CARBOHYD 250 250
FT CARBOHYD 363 363
FT CARBOHYD 705 705
FT CARBOHYD 711 711
FT CARBOHYD 1070 1070
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CR664;

Query Match 79.5%; Score 35; DB 1; Length 1173;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRHL 7
Db 1109 TAYRWHL 1115

RESULT 10
YBP2_STRAU STANDARD; PRT; 150 AA.
ID YBP2_STRAU
AC P29714;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in BPOA2.5' region (Fragment).
OS Streptomyces aureofaciens.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1894;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10762 / CCM 3339;
RX MEDLINE=92407485; PubMed=15274921;
RA Pfeiffer O., Pelletier I., Altenbuchner J., van Pee K.-H.;
RT "Molecular cloning and sequencing of a non-haem bromoperoxidase gene
from Streptomyces aureofaciens ATCC 10762."
RL J. Gen. Microbiol. 138:1123-1131(1992).
CC -1- SIMILARITY: BELONGS TO THE DPS FAMILY.

```

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: M84990; AAB84314.1; -
DR PIR: S27613; S27613.
DR InterPro: IPR002177; DPS.
DR Pfam: PF02047; DPS; 1.
DR ProDom: PD149803; DPS; 1.
DR ProSITE: PS00818; DPS; 1.
DR PROSITE: PS00819; DPS; 2; 1.
DR Hypothetical protein: Plasmid.
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 16648 MW; 45EF8CD9573E875C CRC64;

Query Match
Best Local Similarity 72.7%; Score 32; DB 1; Length 150;
Matches 6; Conservative 1; Mismatches 0; Gaps 0; Indels 0;

OY 1 TAYRYHL 8
DB 64 TAYQHL 71

RESULT 11
OPAF_NEIGO STANDARD; PRT; 234 AA.
AC 004874;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Opacity protein OPA51 precursor (Fragment).
OPAF.
GN Neisseria gonorrhoeae.
OS Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11 / F3;
RC MEDLINE=93178439; PubMed=8440254;
RA Kusch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT "Variable opacity (Opa) outer membrane proteins account for the cell
RT tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT epithelial cells."
RL EMBL J. 12:641-650(1993).
CC -1- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC VARIATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: 218928; CAA79361.1; -
DR PIR: S28628; S28628.
DR PIR: S36329; S36329.
DR InterPro: IPR003394; Porin_opacity.
DR Pfam: PF02462; Opacity; 1.
KW Outer membrane; Multigene family; Signal.
FT NON_TER 1 1
FT SIGNAL <1 1
FT CHAIN 2 >234 POTENTIAL.
FT NON_TER 234 234 OPACITY PROTEIN OPA51.
FT NON_TER 234 234

```

```

SQ SEQUENCE 234 AA; 26772 MW; 9FE5B5DABBA96CA CRC64;

Query Match
Best Local Similarity 72.7%; Score 32; DB 1; Length 234;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYH 6
DB 206 TGYRYH 211

RESULT 12
OPAF_NEIGO STANDARD; PRT; 234 AA.
AC 004878;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Opacity protein OPA55 precursor (Fragment).
OPAF.
GN Neisseria gonorrhoeae.
OS Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS11 / F3;
RC MEDLINE=93178439; PubMed=8440254;
RA Kusch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;
RT "Variable opacity (Opa) outer membrane proteins account for the cell
RT tropisms displayed by Neisseria gonorrhoeae for human leukocytes and
RT epithelial cells."
RL EMBL J. 12:641-650(1993).
CC -1- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA
CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE
CC VARIATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: 218933; CAA79366.1; -
DR PIR: S28632; S28632.
DR InterPro: IPR003394; Porin_opacity.
DR Pfam: PF02462; Opacity; 1.
KW Outer membrane; Multigene family; Signal.
FT NON_TER 1 1
FT SIGNAL <1 1
FT CHAIN 2 >234 POTENTIAL.
FT NON_TER 234 234 OPACITY PROTEIN OPA55.
SQ SEQUENCE 234 AA; 26881 MW; 8EBE30B3A774C766 CRC64;

Query Match
Best Local Similarity 72.7%; Score 32; DB 1; Length 234;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYH 6
DB 206 TGYRYH 211

RESULT 13
OPAF_NEIGO STANDARD; PRT; 234 AA.
AC 004879;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Opacity protein OPA56 precursor (Fragment).

```

GN OPAF.  
 OS Neisseria gonorrhoeae.  
 CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 CC NCBI\_TaxID=485;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN-MS11 / F3;  
 CC MEDLINE=93178439; PubMed=8440254;  
 CC Kupsch E.-M., Knepper B., Kuroki T., Heuer I., Meyer T.F.;  
 CC "Variable opacity (Opa) outer membrane proteins account for the cell  
 CC tropisms displayed by Neisseria gonorrhoeae for human leukocytes and  
 CC epithelial cells."  
 CC EMO J. 12:641-650(1993).  
 CC -1- FUNCTION: IMPLICATED IN A NUMBER OF ADHERENCE FUNCTIONS. OPA  
 CC PROTEINS ARE IMPLICATED IN PATHOGENESIS AND ARE SUBJECT TO PHASE  
 CC VARIATION.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z18934; CAA79367.1; -  
 CC PIR: S28620; S28620.  
 CC InterPro: IPR003394; Portin\_opacity.  
 CC Pfam: PF02462; Opacity; 1.  
 CC Outer membrane; Multigene family; signal.  
 CC FT SIGNAL 1 1 POTENTIAL.  
 CC FT CHAIN 2 >234 OPACITY PROTEIN OPA56.  
 CC FT NON\_TER 234 234  
 CC SEQUENCE 234 AA; 26868 MW; 5175C660839EEFB CRC64;  
 CC  
 CC Query Match 72.7%; Score 32; DB 1; Length 234;  
 CC Best Local Similarity 83.3%; Pred. No. 17;  
 CC Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC 1 TAYRYH 6  
 CC 1  
 CC 206 TGYRYH 211  
 CC  
 CC RESULT 14  
 CC OPRI\_NEIMC STANDARD: PRT; 260 AA.  
 CC AC P10170;  
 CC DT 01-MAR-1988 (Rel. 10, Created)  
 CC DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Opacity-related protein POPM1.  
 CC GN OPR.  
 CC OS Neisseria meningitidis (serogroup C).  
 CC CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 CC CC NCBI\_TaxID=135720;  
 CC CC [1]  
 CC CC SEQUENCE FROM N.A.  
 CC CC STRAIN-C1938 / Serogroup C;  
 CC CC MEDLINE=88260884; PubMed=2455211;  
 CC CC Stern A., Meyer T.F.;  
 CC CC "Common mechanism controlling phase and antigenic variation in  
 CC CC pathogenic neisseriae."  
 CC CC Mol. Microbiol. 1:5-12(1987).  
 CC CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC CC -1- SIMILARITY: STRONG TO THE OPACITY-RELATED PROTEIN POPM3 AND  
 CC CC REGIONS OF HOMOLOG WITH N.GONORRHOEA (STRAIN MS11) OPA GENE  
 CC CC PRODUCTS.  
 CC CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X06445; CAA29748.1; ALT-SEQ.  
 CC PIR: S08514; S08514.  
 CC InterPro: IPR003394; Portin\_opacity.  
 CC Pfam: PF02462; Opacity; 1.  
 CC Outer membrane.  
 CC SEQUENCE 260 AA; 28936 MW; EBA7A2843B3F037B CRC64;  
 CC  
 CC Query Match 72.7%; Score 32; DB 1; Length 260;  
 CC Best Local Similarity 83.3%; Pred. No. 19;  
 CC Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC 1 TAYRYH 6  
 CC 1  
 CC 232 TGYRYH 237  
 CC  
 CC RESULT 15  
 CC OMPC\_NEIGO STANDARD: PRT; 270 AA.  
 CC AC P09888;  
 CC DT 01-MAR-1989 (Rel. 10, Created)  
 CC DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 CC DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 CC DE Outer membrane protein P.IIC precursor (Protein IIC).  
 CC GN PIC.  
 CC OS Neisseria gonorrhoeae.  
 CC CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 CC CC NCBI\_TaxID=485;  
 CC CC [1]  
 CC CC SEQUENCE FROM N.A.  
 CC CC STRAIN-153;  
 CC CC MEDLINE=89096501; PubMed=3145386;  
 CC CC van der Ley P.;  
 CC CC "Three copies of a single protein II-encoding sequence in the genome  
 CC CC of Neisseria gonorrhoeae JS3: evidence for gene conversion and gene  
 CC CC duplication."  
 CC CC Mol. Microbiol. 2:797-806(1988).  
 CC CC -1- FUNCTION: THIS PROTEIN SERVES AS A PORIN.  
 CC CC -1- SUBUNIT: HOMOTRIMER.  
 CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC CC -----  
 CC CC EMBL: X12625; CAA1144.1; -  
 CC CC PIR: S03095; KONH2C.  
 CC CC InterPro: IPR003394; Portin\_opacity.  
 CC CC Pfam: PF02462; Opacity; 1.  
 CC CC Outer membrane; Porin; Transmembrane; Antigen; signal.  
 CC CC FT SIGNAL 1 25  
 CC CC CHAIN 26 270 OUTER MEMBRANE PROTEIN P.IIC.  
 CC CC TRANSMEM 36 44 POTENTIAL.  
 CC CC TRANSMEM 77 85 POTENTIAL.  
 CC CC TRANSMEM 90 96 POTENTIAL.  
 CC CC TRANSMEM 143 157 POTENTIAL.  
 CC CC TRANSMEM 163 173 POTENTIAL.  
 CC CC TRANSMEM 222 234 POTENTIAL.  
 CC CC TRANSMEM 238 246 POTENTIAL.  
 CC CC TRANSMEM 262 270 POTENTIAL.  
 CC CC SEQUENCE 270 AA; 30269 MW; FB8448373830A50D CRC64;  
 CC  
 CC Query Match 72.7%; Score 32; DB 1; Length 270;



Best Local Similarity 83.3%; Pred. No. 20;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAYRYH 6  
| | | | |  
DB 242 TGYRYH 247

Search completed: May 22, 2003, 12:08:40  
Job time : 8.5 secs

GenCore version 5.1.4\_P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 22, 2003, 12:04:33 ; Search time 25 Seconds

(without alignments)  
65.935 Million cell updates/sec

Title: US-09-719-494-12

Perfect score: 44

Sequence: 1 TAYRYHLL 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	81.8	145	16	Q9RT39
2	36	81.8	262	10	Q8W202
3	36	81.8	569	10	Q8W204
4	36	81.8	1199	6	Q28205
5	36	81.8	1372	2	Q54151
6	36	81.8	1373	2	Q9AL58
7	36	81.8	1714	10	Q94HT9
8	36	81.8	1731	10	Q8SB47
9	36	81.8	1733	10	Q94HV6
10	36	81.8	1777	10	Q94HP9
11	36	81.8	1785	10	Q8S789
12	36	81.8	1819	10	Q8SAZ4
13	36	81.8	2162	10	Q9AYC2
14	35	79.5	211	3	Q9HF73
15	35	79.5	323	16	Q8UAD4
16	35	79.5	967	3	Q9Y7H9

17	35	79.5	1064	3	Q9CA00
18	34	77.3	162	16	Q67941
19	34	77.3	361	10	Q94I08
20	33	75.0	72	12	Q98485
21	33	75.0	112	5	Q9U229
22	33	75.0	173	17	Q9YCV1
23	33	75.0	321	16	Q67354
24	33	75.0	374	16	Q9PRX8
25	33	75.0	581	5	Q95T66
26	33	75.0	1766	5	Q9V591
27	32	72.7	71	4	Q9P1E7
28	32	72.7	119	10	Q9FWY3
29	32	72.7	148	2	Q31167
30	32	72.7	172	16	Q8Z542
31	32	72.7	197	5	Q93103
32	32	72.7	214	2	Q51086
33	32	72.7	230	2	Q9R9A7
34	32	72.7	232	2	Q9R9A9
35	32	72.7	232	2	Q9R9A8
36	32	72.7	232	2	Q9K4T9
37	32	72.7	232	2	Q9K4T7
38	32	72.7	232	2	Q9K4T3
39	32	72.7	232	2	Q9R3P5
40	32	72.7	233	2	Q9K4T5
41	32	72.7	234	2	Q07280
42	32	72.7	234	2	Q9R718
43	32	72.7	234	2	Q30759
44	32	72.7	234	2	Q07287
45	32	72.7	234	2	Q9R719

## ALIGNMENTS

## RESULT 1

Q9RT39	PRELIMINARY;	PRF:	145 AA.
AC Q9RT39			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE TRANSPOSASE.			
GN DR1927.			
OS Deinococcus radiodurans.			
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;			
CC Deinococcaceae; Deinococcus.			
OX NCBI_TaxID=1299;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=RI.			
RX MEDLINE=20036896; PubMed=10567266;			
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,			
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,			
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,			
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,			
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,			
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,			
RA Fraser C.M.;			
RT "Genome sequence of the radioresistant bacterium Deinococcus			
RT radiodurans RI."			
RL Science 286:1571-1577(1999).			
DR EMBL; AE002031; AAF11477.1; -.			
DR TIGR; DR1927; -.			
DR InterPro; IPR002686; Transposase_17.			
DR Pfam; PF01797; Transposase_17; 1.			
DR ProDom; PD003831; Transposase_17; 1.			
KW Complete proteome.			
SQ SEQUENCE 145 AA; 16950 MW; 2CD5098D1012CF9 CRC64;			

Query Match 81.8%; Score 36; DB 16; Length 145;  
Best local similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHL 8  
 DB 22 TAYRYHL 29

## RESULT 2

08W202 PRELIMINARY; PRT; 262 AA.  
 AC 08W202;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DE Putative polypeptid.  
 GN OSJNB0028C16.20.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Sasaki C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC098565; AAL69438.1; -  
 KW Polypeptid.  
 SQ SEQUENCE 262 AA; 27461 MW; C7BA920CF2486429 CRC64;

Query Match 81.8%; Score 36; DB 10; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYH 6  
 DB 36 TAYRYH 41

## RESULT 3

08W204 PRELIMINARY; PRT; 569 AA.  
 AC 08W204;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DE Putative polypeptid.  
 GN OSJNB0028C16.18.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
 RA Sasaki C., Henry D., Oates R., Simmons J.;  
 RT "Rice Genomic Sequence."  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC098565; AAL69436.1; -  
 KW Polypeptid.  
 SQ SEQUENCE 569 AA; 60552 MW; 64ED54E8612BE9D7 CRC64;

Query Match 81.8%; Score 36; DB 10; Length 569;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYH 6  
 DB 99 TAYRYH 104

## RESULT 4

028205 PRELIMINARY; PRT; 1199 AA.  
 AC 028205;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE Cofactor D.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96319731; PubMed=8706133;  
 RA Tian G., Huang Y., Rommelaere H., Vandekerckhove J., Ampe C.,  
 RA Cowan N.J.;  
 RT "Pathway leading to correctly folded beta-tubulin."  
 RL Cell 86:287-296(1996).  
 DR EMBL; U61233; AAB17537.1; -  
 SQ SEQUENCE 1199 AA; 133014 MW; 2FCF9C9A972599BA CRC64;

Query Match 81.8%; Score 36; DB 6; Length 1199;  
 Best Local Similarity 85.7%; Pred. No. 16+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 AYRYHL 8  
 DB 985 AYRYHL 991

## RESULT 5

054151 PRELIMINARY; PRT; 1372 AA.  
 AC 054151;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DE HEMAGGLUTININ.  
 GN SHE OR PIC.  
 OS Shigella flexneri, and  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 OX NCBI\_TaxID=623; 562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S.flexneri; STRAIN-SEROTYPE 2A, AND M4243A;  
 RA Noriega F.R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-E.coli; STRAIN=042;  
 RA Henderson I.R., Czechuln J.R., Nataro J.P.;  
 RT "A second autotransporter protease from enteroaggregative Escherichia coli."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U35656; AAB58244.1; -  
 DR EMBL; AF097644; AAD23953.1; -  
 DR MEROPS; S06.005; -  
 DR InterPro; IPR000710; IGA\_S6.  
 DR InterPro; IPR004899; Pertactin\_sup.  
 DR Pfam; PF02395; IGA1; 1.  
 DR Pfam; PF02212; Pertactin; 1.  
 KW Protease; Signal.  
 SQ SEQUENCE 1372 AA; 146449 MW; 68CE71254EAC26EA CRC64;

Query Match 81.8%; Score 36; DB 2; Length 1372;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHL 7  
 DB 1227 TAYRYHL 1233

```

RESULT 6
ID 09AL58 PRELIMINARY; PRT; 1373 AA.
AC 09AL58;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE PIC.
GN Shigella flexneri 2a.
OS Shigella flexneri 2a.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
ON NCBI_TaxID=42897;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YSH60007;
RX MEDLINE=20231776; PubMed=10768931;
RA Al-Hasani K., Henderson T.R., Sakellaris H., Rajakumar K., Grant T.,
RA Nataro J.P., Robins-Browne R., Adler B.;
RT "The sigma gene which is borne on the she pathogenicity island of
RT Shigella flexneri 2a encodes an exported cytopathic protease involved
RT in intestinal fluid accumulation."
RT Infect. Immun. 68:2457-2463(2000).
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN=YSH60007;
RX MEDLINE=21105956; PubMed=11162180;
RA Al-Hasani K., Rajakumar K., Bulach D., Robins-Browne R., Adler B.,
RA Sakellaris H.;
RT "Genetic organization of the she pathogenicity island in Shigella
RT flexneri 2a."
RT Microb. Pathog. 30:1-8(2001).
DR EMBL; AF200692; AK00464.1; -
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertact-sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
SQ SEQUENCE 1373 AA; 146548 MW; 45C016CDA577763 CRC64;

Query Match 81.8%; Score 36; DB 2; Length 1373;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAYRYHL 7
DB 1228 TAYRYHL 1234

RESULT 7
ID 094HT9 PRELIMINARY; PRT; 1714 AA.
AC 094HT9;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative retroelement.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
ON NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPOBARE;
RA Du H., Minx P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,
RA Nascimben L., Preston R., Kirchoff K., King L., Vil M.D., Baker J.,
RA Zlatavarn T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,
RA Cunnius D.M., Ballia V., Shah R., Bahret A., O'Shaughnessy A.,
RA Palmer L., Yang C., Dedhia N., McCombie W.R.;
RT "Genomic Sequence for Oryza sativa, Nippondare strain, clone

```

```

RT OSJNB0036B06, from Chromosome 10, complete sequence."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC051624; AK92560.1; -
DR InterPro: IPR001995; Asprotease_rtrv.
DR InterPro: IPR001969; Asprotease_site.
DR InterPro: IPR005162; Retrotrans_gag.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RYse.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF03732; Retrotrans_gag; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00096; zf_CCHC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN.1.
RW RNA-directed DNA polymerase.
SQ SEQUENCE 1714 AA; 194668 MW; 168C0139B37AF1E0 CRC64;

Query Match 81.8%; Score 36; DB 10; Length 1714;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYH 6
DB 36 TAYRYH 41

RESULT 8
ID 08SB47 PRELIMINARY; PRT; 1731 AA.
AC 08SB47;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative polyprotein.
CN OSJNB0091009.12.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
ON NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence."
RT Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC091732; AAL77161.1; -
RW Polyprotein.
SQ SEQUENCE 1731 AA; 195964 MW; F099666D1D029FB0 CRC64;

Query Match 81.8%; Score 36; DB 10; Length 1731;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TAYRYH 6
DB 103 TAYRYH 108

RESULT 9
ID 094HV6 PRELIMINARY; PRT; 1773 AA.
AC 094HV6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative retroelement.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
ON NCBI_TaxID=4530;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RA Du H., Min P., Abbott A., Doebber A., de la Bastide M., Spiegel L.,
RA Nascimento L., Preston R., Kirchhoff K., King L., Vill M.D., Baker J.,
RA Zutavern T., Santos L., Bell M., Miller B., Kuit K., Rodriguez S.,
RA Cunnius D.M., Balija V., Shah R., Bahret A., O'Shaughnessy A.,
RA Palmer L., Yang C., Dedhia N., McCombie W.R.;
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNB0036806, from Chromosome 10, complete sequence.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC051624; AAK92543.1;
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR005162; Retrotrans_gag.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF003732; Retrotrans_gag; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 1773 AA; 201241 MW; 958C0F0BF77A8AD3 CRC64;

```

```

Query Match
Best Local Similarity 81.8%; Score 36; DB 10; Length 1773;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 TAYRYH 6
DB 99 TAYRYH 104

```

```

RESULT 10
O94HP9 PRELIMINARY; PRT; 1777 AA.
ID O94HP9;
AC O94HP9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative retroelement.
GN OSJNB0004A10.7.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OC NCBI_TaxID=4530;
OX
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RA Nascimento L., Spiegel L., de la Bastide M., Kirchhoff K., King L.,
RA Preston R., Vill M.D., Baker J., Zutavern T., Santos L., Bell M.,
RA Miller B., Kuit K., Rodriguez S., Cunnius D.M., Balija V., Shah R.,
RA Bahret A., Palmer L., Yang C., O'Shaughnessy A., Dedhia N.,
RA McCombie W.R.;
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNB0065C16, from Chromosome 10, complete sequence.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RA McCombie W.R., de la Bastide M., Spiegel L., Nascimento L., Balija V.,
RA Zutavern T., Bell M., Preston R., Kirchhoff K., Kuit K., Baker J.,
RA Santos L., Miller B., Cunnius D.M., Katzenberger F., Muller S.,
RA Shah R., King L., Yang C., Dike S., O'Shaughnessy A., Palmer L.,
RA Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNB0004A10, from Chromosome 10, complete sequence.";
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC074354; AAK92548.1;
DR EMBL; AC098682; AAK01103.1;
DR InterPro; IPR001995; Asprotease_rtrv.

```

```

DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR005162; Retrotrans_gag.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF003732; Retrotrans_gag; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 1777 AA; 201593 MW; CAE2B6BEC53C08E9 CRC64;

```

```

Query Match
Best Local Similarity 81.8%; Score 36; DB 10; Length 1773;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 TAYRYH 6
DB 99 TAYRYH 104

```

```

RESULT 11
O8S789 PRELIMINARY; PRT; 1785 AA.
ID O8S789;
AC O8S789;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative retroelement.
GN OSJNB0093I09.7.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OC NCBI_TaxID=4530;
OX
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RA Spiegel L., de la Bastide M., Kirchhoff K., Preston R., Kuit K.,
RA Nascimento L., Baker J., Vill M.D., Zutavern T., Santos L., Miller B.,
RA Cunnius D.M., Balija V., Shah R., King L., Bell M., Yang C., Dike S.,
RA Palmer L., O'Shaughnessy A., Dedhia N., McCombie W.R.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
RT OSJNB0093I09, complete sequence.";
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC090486; AAK08795.1;
DR EMBL; AC090486; AAK08795.1;
SQ SEQUENCE 1785 AA; 202400 MW; D38E671CAFBA46293 CRC64;

```

```

Query Match
Best Local Similarity 81.8%; Score 36; DB 10; Length 1785;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 TAYRYH 6
DB 99 TAYRYH 104

```

```

RESULT 12
O8SAZ4 PRELIMINARY; PRT; 1819 AA.
ID O8SAZ4;
AC O8SAZ4;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative polyprotein.
GN OSJNB0023P16.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OC NCBI_TaxID=4530;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J., Wilson R., Minx P., Du H.,
RT "Rice Genomic Sequence."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC093568; AAL78107.1; -.
SQ POLYPROTEIN.
KW POLYPEPTIDE.
SEQUENCE 1819 AA; 205815 MW; 6720371C4CFE0BA7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 10; Length 1819;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAYRYH 6
DB 99 TAYRYH 104

RESULT 13
Q9AYC2 PRELIMINARY; PRT; 2162 AA.
ID Q9AYC2;
AC Q9AYC2;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Polyprotein.
GN OSJNB0094H10.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriaristaceae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Spiegel L.A., Nascimento L.V., de la Bastide M., Kirchoff K.A.,
RA King L., Preston R.R., Vill M.D., Baker J.P., Miller B., Zuluaven T.,
RA Rodriguez S., Santos L., Kuit K.H., Cunniff D.M., Balija V.S.,
RA Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA McCombie W.R.;
RT "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT Clone OSJNB0058E19, complete sequence."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC080019; AAK13116.1; -.
DR InterPro: IPR001995; Asprotease_rtrv.
DR InterPro: IPR001969; Asprotease_site.
DR InterPro: IPR005162; Retrotrans_gag.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; Rvise.
DR InterPro: IPR001878; Zn1_CCHC.
DR Pfam: PF03732; Retrotrans_gag; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; rvp; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; ZnF-C2HC; 1.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 2162 AA; 244771 MW; 826F552A5A5C67EF CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 10; Length 2162;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAYRYH 6
DB 483 TAYRYH 488

RESULT 14
Q9HF73 PRELIMINARY; PRT; 211 AA.
ID Q9HF73

```

```

AC Q9HF73;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE Ver153p (Fragment).
OS Ashbya gossypii (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Eremothecaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RA Wendland J., Springer N., Philippson P.;
RT "BEM2 is required for polarized growth and maintenance of cell
RT polarity in the filamentous ascomycete Ashbya gossypii."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF195007; AAG43465.1; -.
FT NON TER 211
SQ SEQUENCE 211 AA; 24538 MW; 74EBF96DC8CE7A1E CRC64;

Query Match
Best Local Similarity 100.0%; Score 35; DB 3; Length 211;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YRYHL 8
DB 118 YRYHL 123

RESULT 15
Q8UD4 PRELIMINARY; PRT; 323 AA.
ID Q8UD4;
AC Q8UD4;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Transcriptional regulator, lysR family.
GN ATU3440 OR AGR_L-2774.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=21608550; Pubmed=11743193;
RX Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayarin T., Levy R., Li M.-J., McClelland E., Palmer A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=11608551; Pubmed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ourullo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Woulam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009274; AAL44253.1; -.
DR EMBL: AE008338; AAK89952.1; -.
KW Complete proteome.
SQ SEQUENCE 323 AA; 35071 MW; BEDBF296CBFFD9F4 CRC64;

```

Query Match 79.5%; Score 35; DB 16; Length 323;  
 Best Local Similarity 87.5%; Pred. NO. 70;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAYRYHLL 8  
 |||||  
 Db 287 TASYRYHLL 294

Search completed: May 22, 2003, 12:09:40  
 Job time : 28 secs